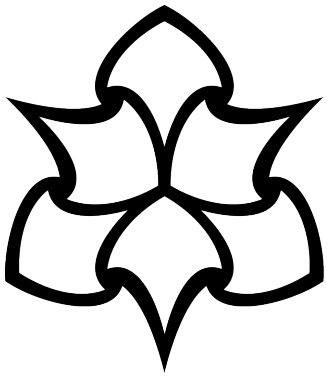


Molecular genetic characteristics of elite rugby union athletes

Shane M Heffernan

September 2016

A thesis submitted in partial fulfilment of the requirement of the Manchester Metropolitan University for the degree of Doctor of Philosophy

Department of Exercise and Sport Science, Sports Genomics Laboratory, Manchester Metropolitan University

Contents

Acknowledgements

Briseann an dúchas trí shúile an lúthchleasaí

(Heredity breaks out in the eyes of the athlete)

Firstly, I would like to thank my director of studies Dr Alun G Williams for his expert advice, guidance and more grammar/spell checking than I would like to admit throughout my PhD journey. On a personal note, it has been an honour, a pleasure and a true privilege learning from, working with and being mentored by an academic of Alun's class. His attention to detail and willingness to go beyond his duties as director of studies are just two of Alun's many attributes that are often underappreciated - *diolch yn fawr iawn* Alun.

Secondly, I would like to extend my particular appreciation to Prof Liam Kilduff for his expert advice in rugby union physiology and his invaluable assistance in athlete recruitment. Without Liam's support the present thesis would not have been possible - *go raibh míle maith agat* Liam. Furthermore, I would like to thank my extended supervisory team, Prof Yannis Pitsiladis, Dr Stephen Day and Dr Jamie McPhee for their support and valuable comments.

Thirdly, I would like to thank the elite athletes and the control participants for their time, DNA and interest in the present research project. Thank you to all the support staff, medical teams, sport scientists, physiologists, coaches and event organisers for their support of the RugbyGene project – again the present thesis would not have been possible without these people.

Finally, I would like to thank my friends and family for their support, encouragement and necessary distractions in recent years. To name a few; my father Michael Gerard Heffernan, my mother Maria Butler, my father's partner Catherine Allen, my adopted uncle Tom Allen, my adopted step brothers Dan and Sean Ellis, my cousin Eimear Butler, my confidante David P Hartley, the mother of my godchild Mairead Aimee Halpin, my academic sounding board Dr Georgina Stebbings, my ever inventive training partner Prof Costis Maganaris and some of my dearest friends Barry O'Mahony, Daniel Hayes, David O'Callaghan, Brendan Brown, Dr Ayser Hussain, Mathew Oliver (and all the MMU Cheshire gym staff) and my PhD, Seeley corridor colleagues over the years - *gratias tibi valde* to you all.

I dedicate this thesis, my entire academic life and the person that I have become to my father Michael Gerard Heffernan. Nothing positive in my life would have ever been possible without your sacrifices, guidance, support and love - I owe you everything.

Publications

The following chapters have been published or submitted for publication;

Chapter 2 **Heffernan S. M.**, Kilduff L. P., Day S. H., Pitsiladis Y. P. & Williams A. G. (2015): Genomics in rugby union: A review and future prospects. *European Journal of Sport Science*, 15(6), 460-468.

Chapter 4 **Heffernan, S. M.**, Kilduff, L. P., Erskine, R. M., Day, S. H., McPhee, J. S., McMahon, G. E., ... & Williams A. W. (2016). Association of *ACTN3* R577X but not *ACE* I/D gene variants with elite rugby union player status and playing position. *Physiological Genomics*, 48(3), 196-201.

Chapter 5 **Heffernan, S. M.**, G. K. Stebbings, L. P. Kilduff, R. M. Erskine, S. H. Day, C. I. Morse, … & A. G. Williams. (2016). Fat mass and obesity associated (*FTO*) gene influences skeletal muscle phenotypes in non-resistance trained males and elite rugby playing position. Submitted to BMC Genetics on 12/08/16.

Chapter 6 **Heffernan, S. M.**, L. P. Kilduff, R. M. Erskine, S. H. Day, J, G. K. Stebbings, C. J. Cook, … & A. G. Williams. (2016). Apolipoprotein E ε4 allele is not associated with elite rugby status but is present in 30% of athletes. Submitted Plos One on 26/08/16.

Chapter 7 **Heffernan, S. M.**, L. P. Kilduff, R. M. Erskine, S. H. Day, J, G. K. Stebbings, C. J. Cook, … & A. G. Williams. (2016). *COL5A1* gene variants previously associated with reduced soft tissue injury risk are associated with elite athlete status in rugby. Submitted to BMC Genomics on 16/09/16.

Conference Proceedings

Heffernan, S. M., Kilduff, L. P., M., Erskine R., Day, S. H., McMahon, G. E., Cook, C. J., … & Williams, A. G. (2014). *ACE* and *ACTN3* gene variants are associated with elite status and playing position in rugby union. In Bases 2013 abstracts. *Journal of Sports Sciences, 32(Suppl 1)*, s49-s50.

Publications associated with this thesis

Bradburn, S., McPhee, J. S., Williams, A., **Heffernan, S. M.**, Lockey, S., Day, S., & Murgatroyd, C. (2015). Protocol for Multiplex Amplicon Sequencing Using Barcoded Primers. In Handbook of Vascular Biology Techniques (pp. 427-438). *Springer* Netherlands.

Williams A. G., Day S. H., Lockey S. J., **Heffernan S. M.** & Erskine R. M. (2014). Genomics as a practical tool in sport - have we reached the starting line? *Cellular and Molecular Exercise Physiology,* 3(1): e6.

Williams A. G., **Heffernan S. M.**, & Day S. H. (2014). Genetic testing in exercise and sport - have direct-to-consumer genetic tests come of age? *Science and Sport: Current Trends*, 2(1), 3-9.

Related conference Proceedings

Lockey S.J., Erskine R.M., Pedlar C., Kipps C.W.G., **Heffernan S.M.**, Herbert A.J., … & Day, S.H. (2015). The individual and combined association between multiple gene polymorphisms and elite endurance athlete status. *European Congress of Sports Science, 2015 Conference.*

Neale, J. P. H., **Heffernan, S. M.**, Day S. H., Erskine R. M. and Williams, A. G. (2015). *ACE* I/D and *ACTN3* R577X genotypes in elite rugby league. In *BASES student* conference 2015.

Lockey S. J., Erskine R. M., **Heffernan S. M.**, Herbert A. J., Pedlar C. G. G., Kipps C. W., … & Day S.H. (2014). *ACTN3 R577X* genotype is not associated with elite marathon runners. *Physiological Society, Physiology 2014 Conference.*

Figures list

Tables list

Table 13 Genotype score details for optimal RU Forwards (GS_F) based on evidence from thesis observations (genotype score chosen from the data contained within the experimental chapters, 4-7).

Table 14 Genotype score details for optimal RU Backs (GS_B) based on evidence from thesis observations (genotype score chosen from the data contained within the experimental chapters, 4-7).

Abbreviations

3′ UTR	3′ untranslated region
AAMP	anaerobic alactic muscular power
ACE	angiotensin converting enzyme
ACL	anterior cruciate ligament
ACTN3	alpha-actinin-3
APOE	apolipoprotein E
ApoE	apolipoprotein E-based peptide
Arg	argenine
BBB	blood brain barrier
BH	Benjamini-Hochberg corrections
BMI	body mass index
BMP4	bone morphogenetic protein 4
bp	base pair
CCL2	chemokine (C-C motif) ligand 2
CI	confidence interval
COL5A1	collagen type V-α-1
COL5A2	collagen type V-α-2
CTE	chronic traumatic encephalopathy
CypA	controlling cyclophilin A
D	deletion
DZ	dizygotic
EAMC	exercise-associated muscle cramping
EE	effect estimate
EI	exertion index
ELN	elastin
F	female
FDR	false discovery rate
FFM	fat free mass
FGF10	fibroblast growth factor 10
FGF3	fibroblast growth factor 3
FGFR1	fibroblast growth factor 1
FS	frozen shoulder
FTO	fat mass and obesity associated gene
FWER	family-wise error rate
GNB3	guanine nucleotide-binding protein
GPS	global positioning system
GS	genotype score
GWAS	genome-wide association studies
HW	Hardy-Weinberg equilibrium
HWP	Hardy-Weinberg principle
I	insertion
IGF-1	Insulin-like growth factor 1
IGF2	Insulin-like growth factor 2
IL-6	interleukin 6
IL-12	interleukin 12
IRX3	iroquois homeobox 3
JNK	c-Jun N-terminal kinases
LM	lean mass
M	male
MM	muscle mass
MC	multiple test correction

MSTN	myostatin
mTBI	mild traumatic brain injury
MTC1	Mtc1p
Myf5	myogenic factor 5
MZ	monozygotic
NCSMTI	non-contact soft musculoskeletal tissue injuries
NS	not stated
OLIG2	Oligodendrocyte transcription factor 2
OR	odds ratio
PPARA	peroxisome proliferator-activated receptor α
PPARGC1A	peroxisome proliferator-activated receptor gamma coactivator 1-alpha
RFU	Rugby Football Union
RHIE	repeated high intensity exercise
RL	rugby league
ROC	receiver operating characteristic
RU	rugby union
Sig test	significance test
SMCP	single muscular contraction power
SNP	single nucleotide polymorphisms
TBI	traumatic brain injury
TE	tennis elbow
TGS	total genotype score
TNF	tumour necrosis factor
TNF-α	tumour necrosis factor alpha
UCP2	uncoupling protein 2
VDR	vitamin D receptor
WGS	weighted genotype score

Abstract

The capacity to achieve elite athletic success is known to be highly heritable and physiologically complex. Recently, there has been a substantial rise in the number of peer reviewed sports genomics publications, however the majority of these have focused on the genetic components of either strength/sprint/power or endurance athletes, with little attention given to team sports. Since rugby union (RU) athletes perform under a well-defined set of rules and parameters, which are ubiquitous across all playing positions, they present an ideal cohort via which to study the importance of genetic variation in sport and as such is the purpose of the present thesis. The aims of the present thesis were, firstly, to recruit a large cohort of elite RU athletes and compare the genetic profile of these athletes to that of a control population. Secondly, because of the large physiological differences between elite RU playing position, the present thesis further aimed to compare the genetic diversity in RU inter-positional variation. 1164 participants, consisting of 454 elite RU athletes and 710 controls from the general population were recruited for the present thesis. Genotyping data was generated for SNPs within the *APOE* ε4 (rs429358 and rs7412), *ACTN3* (rs1815739), *ACE I/D* or *ACE* tag SNP (rs4341), *COL5A1 (*rs12722 and rs3196378) and *FTO* (rs9939609) genes. Only the *COL5A1* SNPs were different when comparing the entire elite RU cohort to controls. However, observations of *APOE* ε4 frequency did reveal that ~30% of elite RU athletes are at higher risk of poor outcome following mTBI (concussion). Regarding playing position, both *ACTN3* and *FTO* showed position specific variation, with the greatest associations in the backs playing position. There were no associations between elite RU athletes and controls for *ACE* I/D. When incorporating all of the above SNPs into a polygenic profile, the entire elite RU cohort, the backs and forwards showed significant deviation from controls. Interestingly by using an ROC model the present thesis identified significant polygenic sensitivity in discriminating between elite RU backs and forwards which trended towards the backs. The data generated in the present thesis are the first to show genetic variation in a large cohort of elite RU athletes, with particular emphasis on positional

specificity. These data are only the ‘tip of the iceberg’ in understanding the molecular aspect of elite RU physiological underpinnings and analysis of many more variants are required in addition to replication of the present results. Nonetheless, these data are the first step in this understanding and may have implications in positional selection, position specific training and injury management, in the future. Follow up studies should focus on further recruitment of elite RU athletes and combining genetic data with phenotypic data, specific to elite RU athletes and with particular attention to injury susceptibility.

Chapter 1

Thesis introduction

1.1 Introduction

The majority of scientific investigation into player performance in rugby union has focused on environmental factors such as training methods, dietary supplementation and recovery strategies (West *et al.*, 2013; Barr *et al.*, 2014; Bradley *et al.*, 2015), with a great deal of epidemiological research also investigating injury frequency and risk including a recent focus on brain injury (Raftery, 2013; Gardner *et al.*, 2014; Fuller *et al.*, 2015a; Williams *et al.*, 2015b; Fuller *et al.*, 2016). However, considerable evidence shows that performance and injury traits are highly (but variably) heritable (Simoneau & Bouchard, 1995; Bouchard *et al.*, 1999; Hakim *et al.*, 2003; De Moor *et al.*, 2007; Peeters *et al.*, 2007), yet little scientific effort has been invested to elucidate this inter-individual variation within rugby union. For example, Simoneau and Bouchard (1995) showed the genetic heritability of muscle fibre type proportion was ~50%, with larger heritability for body mass index and mass at ~60%, height at ~80%, elbow flexion and knee extension strength at ~50% and hand grip strength at ~60% (Silventoinen *et al.*, 2008).

Rugby union athletes are qualitatively and quantitatively different from other athlete groups (both individual and team) in that there are vast differences in the physiological and anthropometric characteristics across a single rugby union team according to playing position (Smart *et al.*, 2013). Rugby union is also distinctive as individual clusters of positions require different movement patterns in elite players (Quarrie *et al.*, 2013) and thus differ in their metabolic demands. For example, anthropometric and physiological variables differ significantly across playing positions, with elite scum half players averaging ~177 cm and ~85 kg, in contrast to props averaging ~185 cm and ~117 kg – a difference of 8 cm and 32 kg (Fuller *et al.*, 2013). Furthermore, in terms of positional specific physiological differences that may be reflected in players' genetic variation, backs show lower maximal strength compared to forwards in terms of bench press (difference ~11 kg), back squat

(difference ~18 kg) and power clean (difference ~9 kg; Smart et al., 2014). However, backs are faster, sprinting 10 m (difference = ~0.09 s) and 20 m (difference = ~0.11 s) than forwards (Smart *et al.*, 2014) and these differences become larger when specific positions are considered (Smart *et al.*, 2013). Positional differences are further evidenced by game demand data that shows the requirement for differing metabolic capacities dependent on playing position (Deutsch *et al.*, 2007; Roberts *et al.*, 2008; Cahill *et al.*, 2013; Quarrie *et al.*, 2013). For example, backs travelled 12% greater total distance (6545 m versus 5850 m), achieved maximum speeds 16% faster (30.4 $km \cdot h^{-1}$ versus 26.3 $km \cdot h^{-1}$) and engaged in over four times more (58% versus 13%) high-intensity running activities (> 5.0 $m \cdot s^{-1}$), as a proportion of total activity compared to forwards (Roberts *et al.*, 2008; Cahill *et al.*, 2013).

In sports genomics, elite sprint/power athletes are often compared to elite endurance athletes for the purpose of elucidating genetic variation at the ends of the human physical performance spectrum (Eynon *et al.*, 2009; Eynon *et al.*, 2010; Eynon *et al.*, 2011b; Ruiz *et al.*, 2011a; Eynon *et al.*, 2013b). However, these athletes often originate from different geographical regions where socioeconomic backgrounds and levels of sporting professionalism in their community differ and these things are likely to drive their choice to engauge with a particular sporting discipline, not their inate physiological ability. These considerations are less of a challenge in the elite rugby union community, particularly when considering the genetic variation within playing position, as individuals choose to play rugby prior to being selected into their most suitible playing position. As such, since rugby union athletes perform under a well-defined set of rules and parameters, they present an ideal cohort via which to study the importance of genetic variation in sport and thus is the purpose of this thesis.

1.1.1 Thesis overview

Chapter *2.1* outlines the current knowledge of rugby union characteristics. Part one will discuss positional diversity within rugby union athletes considering anthropometric and physiological phenotypic data. Game monitoring technology such as time motion analysis and global positioning system (GPS) systems have allowed us to estimate the metabolic demands of match play and are widely used, in the field, by practical scientists (Twist & Worsfold, 2015). Thus, to assess the molecular genetic differences in playing position, reviewing and extrapolating the appropriate data in relation to these metabolic demands (i.e. different activation of molecular pathway proportions) are vital. Furthermore, given the collision and high speed nature of rugby union, injuries are a common and ever growing issue within the game. World Rugby and national governing bodies (such as the Rugby Football Union (RFU)) have invested substantial resources in assessing the current state of injuries within the elite game. Understanding the genetic underpinnings of injury susceptibility may aid in injury management in the future and this thesis embarks on the first step to this potential. Given the current interest in concussion (Raftery, 2013) and the known clinical impact of such injuries (Bennett *et al.*, 2013), this particularly important aspect of injury is reviewed separately.

Chapter *2.2* reviews and discusses the current knowledge of physical performance trait heritability. It is generally accepted that the physiological/psycho-physiological components of any human performance trait is a composite of environmental (infant nutritional status, socioeconomic considerations, training etc.) and heritability (genetics). As such, attempting to study any human performance trait without acknowledging the heritable element renders any discussion incomplete and these topics are therefore reviewed herein. The study of sports genomics has grown substantially in recent years (Ahmetov & Fedotovskaya, 2012), with larger athlete biobanks being generated worldwide (Pitsiladis *et al.*, 2013) and coming

together in the Athlome consortium (Pitsiladis *et al.*, 2016). Even though the field shows considerable promise in understanding the molecular basis of elite athlete physiology, in relation to team sport athletes, few investigations have embarked upon understanding the genetic variation. Those that have, often show considerable methodological flaws and are reviewed (Chapter *2.2.4*), with the intention of addressing these concerns in the present thesis. An inherited proportion of rugby union ability was proposed in the early 20th century (Jack, 1922) with further small pilot investigations showing potential for genetic variation (Bell *et al.*, 2009; Bell *et al.*, 2010; Bell *et al.*, 2012c). These primary inquiries are thoroughly reviewed herein. Following a broad review and introduction into the field of sports genomics, a specific detailed review of the candidate genes that were investigated within this thesis and their rationale for inclusion are discussed (Chapter *2.3*). These review elements consist of the fundamental rationale generating the aims of the thesis.

Because of the similarities in the scientific methods used for the experimental chapters (*4*, *5*, *6* and *7*), chapter *three* presents a detailed description of the methodology applied and will be referred to as the experimental chapters unfold. The following five chapters consist of the main body of the thesis and the purpose of its production - the experimental chapters (*4*, *5*, *6*, *7* and *8*). Here, all experimental data are presented in journal article format, link directly to the present thesis aims and contribute to the further understanding elite rugby union molecular physiology. Chapter *nine*, the general discussion, amalgamates the thesis components, explores the findings in a deeper context and considers the limitations, future direction of research and possible practical implications of the molecular underpinnings inherent to elite rugby union physiology.

Chapter 2

Literature review

A portion of part *2.2* of this chapter is published in:

Heffernan S. M., Kilduff L. P., Day S. H., Pitsiladis Y. P. & Williams A. G. (2015). Genomics in rugby union: A review and future prospects. *European Journal of Sport Science*, 15(6), 460-468.

2.1 Part 1: Rugby union athlete characteristics

2.1.1 Elite player physiological characteristics

Rugby union is an intermittent team sport that requires a high level of strength, power, speed and aerobic/anaerobic fitness (Roberts *et al.*, 2008; Twist & Worsfold, 2015). Compared to most other team sports that require these traits, rugby is unique, as it includes large acceleration forces often culminating in high impact collusions (Cunniffe *et al.*, 2009; Owen *et al.*, 2014; Bradley *et al.*, 2015). Furthermore, rugby union is comprised of diverse playing positions - forwards (loose-head props, hookers, tight-head props, left locks, right locks, blind-side flankers, open-side flankers, and number eights) and backs (scrum halves, fly halves, left wings, inside centres, outside centres, right wings, and full backs). Each have different physiological and technical attributes (Deutsch *et al.*, 2007; Roberts *et al.*, 2008; Cahill *et al.*, 2013; Quarrie *et al.*, 2013). Briefly, along with the hooker, the loose-head and tight-head props make up the front row forwards, which refers to their position in the scrum (appendix 4). To be successful, both props and the hooker must be extremely strong in the neck, shoulders, upper body and legs. Locks make up the second row forwards and are described as the 'power house' of the scrum. They are required to be tall, heavy and powerful players. Flankers and number eights must have considerable speed, strength, and fitness attributes. Scrum half's tend to be smaller players, act as the link between the forwards and the backs and needs good vision, speed and awareness. Fly half's are the 'play makers', responsible for coordinating the attack and defiance, while being physically capable of exploiting weaknesses in the opposition. Wingers are generally the fastest players that require great relative power to finish planned moves and attacks. Centres tend to be strong and powerful athletes, dynamic runners that again exploiting weaknesses in the opposition. Fullbacks are similar to wingers in that they are among fastest players that require great relative power to finish planned moves and attacks. Additionally, fullbacks have a greater

responsibility in defiance and are often required to make high impact tackles (England Rugby, 2016). For a more detailed description of the player roles, see appendix 4.

Rugby athletes are required to perform "static and dynamic exertion efforts" (scrums, mauls, rucks and lineouts) when competing to gain or maintain possession of the ball (Deutsch *et al.*, 2007; Smart *et al.*, 2014), which requires athletes to develop and maintain high force and velocity dominant physiological phenotypes (Cross *et al.*, 2014; Table 1). However, these and other physiological abilities are highly positional specific and require particular inherent abilities to perform a given static and dynamic exertion effort. For example height, in combination with their lower body power output and vertical jump performance, is vital for a forwards to be successful in rugby union lineouts (Table 1). Evidence now supports the notion that since the advent of professionalism in rugby union, over 20 years ago, players have become progressively taller, heavier, stronger and faster (Appleby *et al.*, 2012; Sedeaud *et al.*, 2012; Barr *et al.*, 2014). From the assessment of Rugby World Cup data (1987 to 2007), forwards and backs have become heavier by ~6.6 kg and ~6.7 kg and taller by ~0.6 cm and ~1.1 cm, respectively, with the most successful teams consisting of the tallest backs and heaviest forwards (Sedeaud *et al.*, 2012). Furthermore, the positional-specific physical evolution of elite rugby athletes remains consistent in both European and Southern hemisphere domestic competition (Fuller *et al.*, 2013; Sedeaud *et al.*, 2013; Smart *et al.*, 2013). Rugby athletes are generally taller, heavier and leaner (Table 1) than equivalent athletes prior to the professional era (Sedeaud *et al.*, 2013). This suggests that by artificial selection, coaches now favour more physically imposing athletes (Table 1), which will be partly determined by genetic inheritance

Table 1 Rugby union (RU) players and general population anthropometric and physiological descriptive variables presented as means. Some rugby union variables presented as positional subgroups. Dashed cells (——) represents an absence of data within the available literature.

Phenotype	Group									
	General population	**Forwards**	**Backs**	**Props**	**Locks**	**Hookers**	**Backrow**	**Half backs**	**Centres**	**Back 3**
Anthropometrics										
Height (m)	1.78	1.89	1.82	1.85	1.98	1.81	1.90	1.79	1.84	1.83
Mass (kg)	75	111	91	117	113	104	106	87	96	92
Body fat (%)	22	13	10	16	13	15	12	11	11	9
FFM (kg)	64	*92	*81	94	95	89	90	79	84	81
Fibre type (% of type II)	53	53	56	_____	_____	_____	_____	_____	_____	_____
Strength and Power										
Lower body power output				(Props, Locks and Hookers)						
• **W·kg^{-1}**	11	*51	*58		48		54	56	58	60
• **W**	221	*5606	*5370		5416		5796	5000	5680	5431
Max bench press (kg)	57	136	125	133	121	124	119	111	113	109
Max back squat (kg)	69	186	168	184	157	185	169	155	163	157
Max power clean (kg)	_____	104	95	102	102	101	98	91	93	91
Running speed										
Max sprint times										
• **10 m**	1.90	1.78	1.69	1.85	1.79	1.81	1.76	1.72	1.70	1.68
• **20 m**	3.25	3.07	2.96	3.21	3.13	3.14	3.06	2.96	2.95	2.89
• **30 m**	4.42	_____	4.04	_____	_____	_____	_____	4.14	4.12	4.11
Game Demands										
Total Distance covered (m)	_____	4679	5957	3698	5027	4746	5244	5693	5907	6272
Distance covered at (> 5 m·s^{-1}) high speed (m)	_____	178	511	102	158	147	306	381	586	566
High-intensity static exertion activities (min:s)	_____	7:56	1:18	(Props and Locks) 8:03		(Backrow and Hookers) 7:47		(Out halves and Centres) 1:33		1:05
RHIE				(Props, Locks and Hookers)						
• **Bouts**	_____	*12	*6		11		13	5	7	6
• **Recovery from bouts (s)**		*428	*638		398		457	612	751	551

Adapted from (Jardine *et al.*, 1988; George *et al.*, 1999; Scott *et al.*, 2003; Glaister *et al.*, 2008; Roberts *et al.*, 2008; Crewther *et al.*, 2009; Nuzzo *et al.*, 2010; Ahmetov *et al.*, 2011; Menzel & Hilberg, 2011; Santiago *et al.*, 2011; Chtourou *et al.*, 2012; Crewther *et al.*, 2012; Fuller *et al.*, 2013; Smart *et al.*, 2013; Smart *et al.*, 2014; Jones *et al.*, 2015).

FFM, fat free mass; RHIE, repeated high intensity exercise. *Value generated by averaging the individual playing position data in the absence the data from the literature.

11 | P a g e

2.1.2 Anthropometrics

Given the positional selective pressures and performance roles, backs (generally speaking) are shorter (~0.07 cm), have less mass (~10 kg), less FFM (~11 kg) and a higher proportion of type II muscle fibres (~3%) than forwards (Jardine *et al.*, 1988; Smart *et al.*, 2013; Smart *et al.*, 2014). These anthropometric quantities are indicative of the body type required for backs, specifically the back three (Table 1), to perform high velocity running – the main performance criterion of their selection to elite status (Chapter 2.1.3). Correspondingly, the opposite anthropometric values (i.e. taller, heavier etc.) show the required body type for forward positional requirements (Chapter 2.1.3).

2.1.3 Strength and Power

Arguably, strength and power capacities better convey the positional variation within rugby union. Backs show lower maximal strength compared to forwards in terms of bench press (~11 kg), back squat (~18 kg) and power clean (~9 kg). However, backs are faster sprinters over 10 m (~0.09 s) and 20 m (~0.11 s) than forwards (Smart *et al.*, 2014; Table 1) and show considerable variation across rugby union playing positions. Importantly, note the comparisons between absolute/relative peak power outputs and playing position (Table 1). Considering relative power output, as a ratio scale of mass, the back positions have greater values ($W \cdot kg^{-1}$) than forwards, however the relationship is reversed when considering absolute lower body power output (Crewther *et al.*, 2012). This scaling variation is an important consideration for accurate field based physiological testing but also in assessing molecular genetic characteristics of muscle phenotypes because of the individuality of genetic variation, such as fat and muscle mass.

2.1.4 Running speed

One of the most vital physiological quantities for backs, particularly back three and centres, is the ability to achieve high sprinting velocities. Backs achieve faster running velocities

over 10 m (~0.09 s), 20 m (~0.10 s) and this positional variation exemplifies considering specific playing positions (Table 1). Importantly in the context of case-control genetic association studies, such as the present thesis, there are considerable physiological differences from the general public (Table 1). Props have only 0.04 s faster 20 m sprint capacity to the general population, whereas the back three players are 0.36 s faster (Table 1). All rugby athletes undergo rigorous training regimes for speed (Barr *et al.*, 2014), yet there is such a slight difference between props and the general population. This difference is likely to be an artefact of both specific training and the genetic determinant of trainability, where a response to a particular training modes are modulated by genotypic and allelic variation (Bouchard *et al.*, 1999). It is highly likely that prior to elite selection, young athletes were at least partially selected to their playing position based on their affinity to a particular physiological capacity (strength, mass, sprint, repeated sprint recovery etc.).

2.1.5 Game demands

Recently, GPS tracking and time-motion analysis have been used to estimate the physical demands of rugby athletes and compare forwards to backs during high-level match play (Roberts *et al.*, 2008; Cahill *et al.*, 2013; Johnston *et al.*, 2014; Jones *et al.*, 2015). Backs travel 12% greater total distance (6545 m versus 5850 m), achieved maximum speeds 16% faster (30.4 $km \cdot h^{-1}$ versus 26.3 $km \cdot h^{-1}$) and engaged in over four times (58% versus 13%) high-intensity running activities (> 5.0 $m \cdot s^{-1}$), as a proportion of total activity compared to forwards (Roberts *et al.*, 2008; Cahill *et al.*, 2013). These data suggests a more sprint-oriented metabolic demand in backs compared to forwards. Furthermore, due to the complexities of forward play, forwards performed sixfold more (9.9%) high-intensity static exertion activities (rucks, mauls, scrums and line-outs) than backs (1.6%) and spend 19.8% more time running above 80% of their maximal speed (Roberts *et al.*, 2008; Cahill *et al.*, 2013, respectively). This implies that forwards, although often of higher body mass, (Fuller

et al., 2013) are more likely to benefit from fatigue-resistant physiological qualities than backs. Accordingly, Deutsch *et al* (2007) showed that forwards had a notably higher work-to-rest ratio than backs (1:7 and 1:22, respectively) in UK top flight Northern hemisphere rugby. Furthermore, in Sothern hemisphere top flight competition, lower work-to-rest (1:4 and 1:6) were identified but showing a similar relationship to the UK data (Figure 1; Austin *et al.*, 2011a). Given that the roles of backs and forwards differ significantly in terms of physiological demands, these differences may be reflected in distinct genetic characteristics.

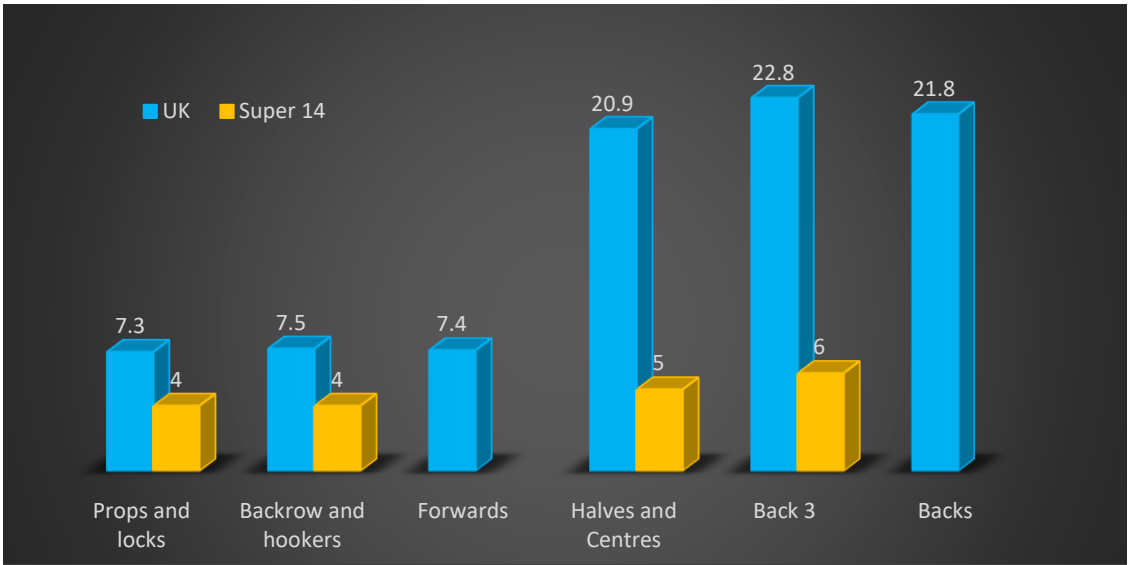


Figure 1 Work-to-rest data by positional group for top tier UK and Super 14 rugby union. Adapted from Austin et al. (2011a) and Deutsch et al. (2007). Chart numbers are the units of work as a ratio of one unit of rest (1:X).

More recent technological advances in game analysis have seen the inclusion of accelerometer data in addition to GPS allowing practitioners to more stringently assess repeated high-intensity exercise (RHIE) and collisions (Jones *et al.*, 2015; Twist & Worsfold, 2015). A RHIE bout is defined as, > 3 high accelerations (> 2.79 $m \cdot s^{-1}$), high speed (>5 $m \cdot s^{-1}$) or contact efforts with less than 21 s recovery between bouts (Spencer *et al.*, 2004; Austin *et al.*, 2011b; Gabbett *et al.*, 2012). As before, significant positional differences exist (Table 1) under these defined parameters (hookers and backrow = 13 bouts versus half backs = 5 bouts) and are further exemplified by the recovery time between bouts

(front five = 398 s, halves = 612 s; Jones *et al.*, 2015). As such, these collective movement patterns suggests a requirement for a high proportion of fast twitch muscle fibres and highly developed anaerobic energy metabolism within the backs, with the forwards' patterns suggesting a requirement for a high proportion of slow twitch fibres, greater capacity for rapid recovery between high-intensity static exertion/RHIE activities and a more developed aerobic endurance capacity. Given the highly heritable nature of these and other physical traits, and their contribution to success in a given playing position, these physiological differences should be reflected in distinct genetic characteristics.

Furthermore, these recent technologies allow for the quantification of match play collisions and unsurprisingly forwards perform more game impacts than backs (Cunniffe *et al.*, 2009; Bradley *et al.*, 2015; Owen *et al.*, 2015). In fact, when considering only the highest level impacts, forwards were involved in 60% more collisions (Cunniffe *et al.*, 2009). Moreover, front row and scrum half athletes showed greater low-intensity impacts, with backrow players experiencing the greatest number of high-intensity impacts (Owen *et al.*, 2015). These data present an insight into the true physiological demands of rugby union, given its collision nature (Twist & Worsfold, 2015). Importantly, Jones et al. (2014) recently showed the relationship between the change in creatine kinase levels at both +16 h ($r = 0.438$) and +40 h ($r = 0.638$) and the number of game impacts. Combining these data with the elite rugby work-to-rest data (Figure 1), the heritable notion of artificial selection in assigning playing position is further supported.

2.1.6 Injury epidemiology

Owing to the collision nature of rugby union, injuries are a common and a growing artefact of the increasing size and strength of the athletes (Cunniffe *et al.*, 2009; Sedeaud *et al.*, 2013; Owen *et al.*, 2014; Bradley *et al.*, 2015) and thus greater changes in momentum during player

collisions, as well as during voluntary accelerations and decelerations. This has, resulted in rugby union having one of the highest reported injury incidence in professional team sports (Brooks & Kemp, 2008). In recent years World Rugby and the RFU have instigated injury surveillance schemes in order to quantify the incidence and severity of injuries at the elite level of the game (Fuller *et al.*, 2008; England Professional Rugby Injury Surveillance Project Steering Group, 2016). The greatest achievement here is the standardised method of data collection and analysis that can be applied across all rugby codes (Fuller *et al.*, 2007) and allows for pooling of data from many teams and competitions to give an accurate picture of the injury epidemiology (Fuller *et al.*, 2015a). Indeed, injury assessment is not only important for player safety, but also results in rule changes (Trewartha *et al.*, 2015). Player injury results in an inability to select the best players and disruption to an athletes training availability. As such, recent data shows a negative correlation between days absence through injury (22% of mean) and team success, as defined by competitive league points tally (Figure 2; Williams *et al.*, 2015b). Using genetic data for injury susceptibility could enhance individualised player management, in that training load and game time could be adjusted to allow increased availability for matches and greater team success.

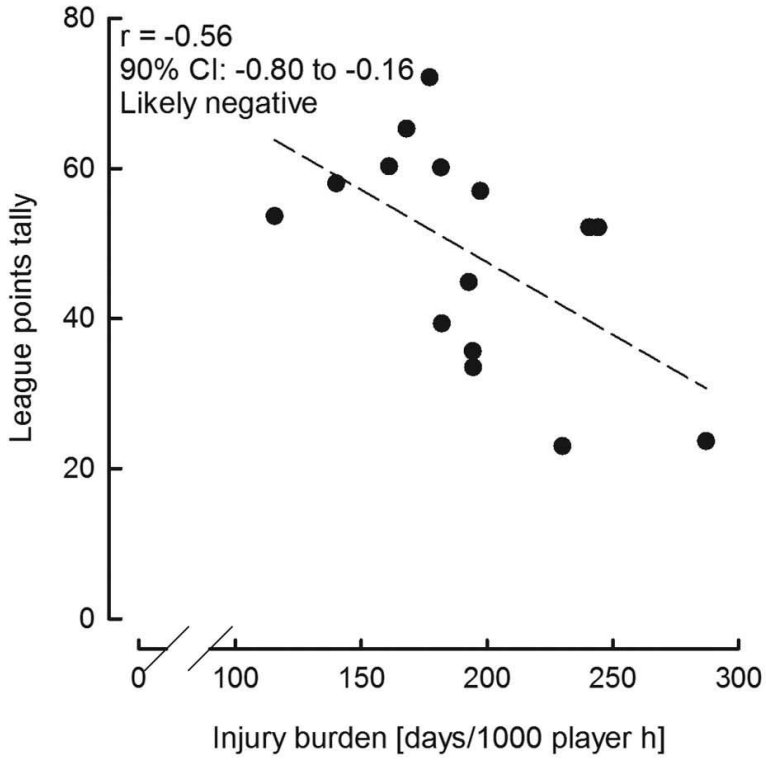


Figure 2 Pearson correlation, 90% CI and qualitative inference for team-averaged values of each injury (injury burden) and team success (league points tally) outcome. Smallest worthwhile effect: r ± 0.3. Adapted from Williams et al. (2015b).

Meta-analysis data have shown that for every 1000 h of match play, a player will typically experience 81 injuries and three per 1000 h from training (Williams *et al.*, 2013). However, no difference between backs and forwards was identified (Williams *et al.*, 2013). Interestingly, figures 3 and 4 show that the majority of these injuries are ligament, tendon and muscle injuries of the lower limb (Williams *et al.*, 2013). This is not surprising as the majority of injuries occur during the tackle (Figure 5), which tends to involve lower body collisions. These data are supported by more recent results from Southern hemisphere Super Rugby, where injury incidence per 1000 h was 83 for game play and two per 1000 h during training (Schwellnus *et al.*, 2014). Williams et al. defined injury severity as minimal (2-3 days), mild (4-7 days), moderate (8-28 days) and severe (> 28 days). Of these definitions, the most common (moderate) resulted in 28 incidence per 1000 h followed by mild (23 incidence), minimal (17 incidence) and severe (15 per 1000 h; Williams *et al.*, 2013). Adding genetic data to a player's injury and performance profile might aid in athlete welfare and injury management.

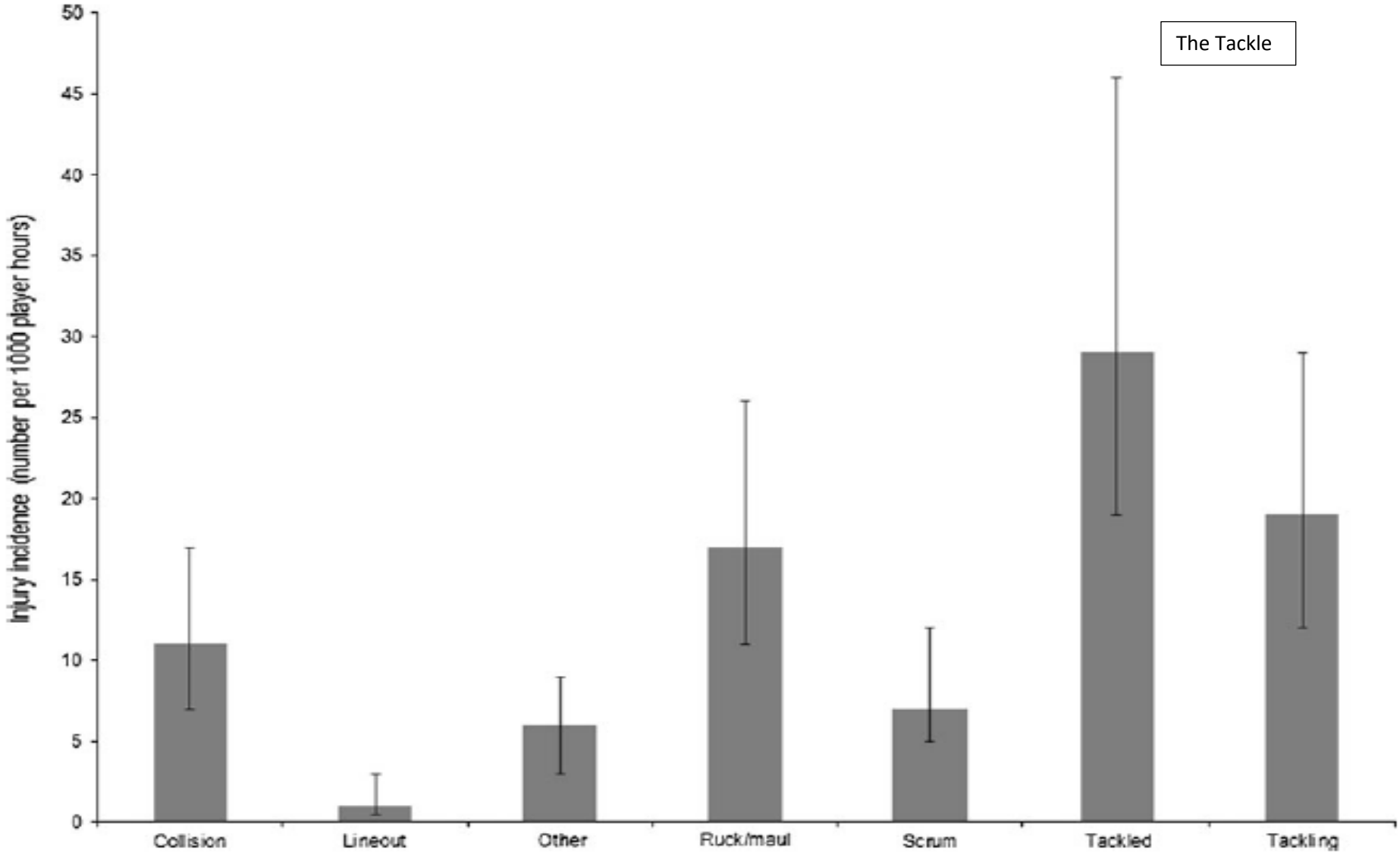


Figure 3. Injury incidence (95 % CI) by cause of injury incident. Adapted from Williams et al., (2013).

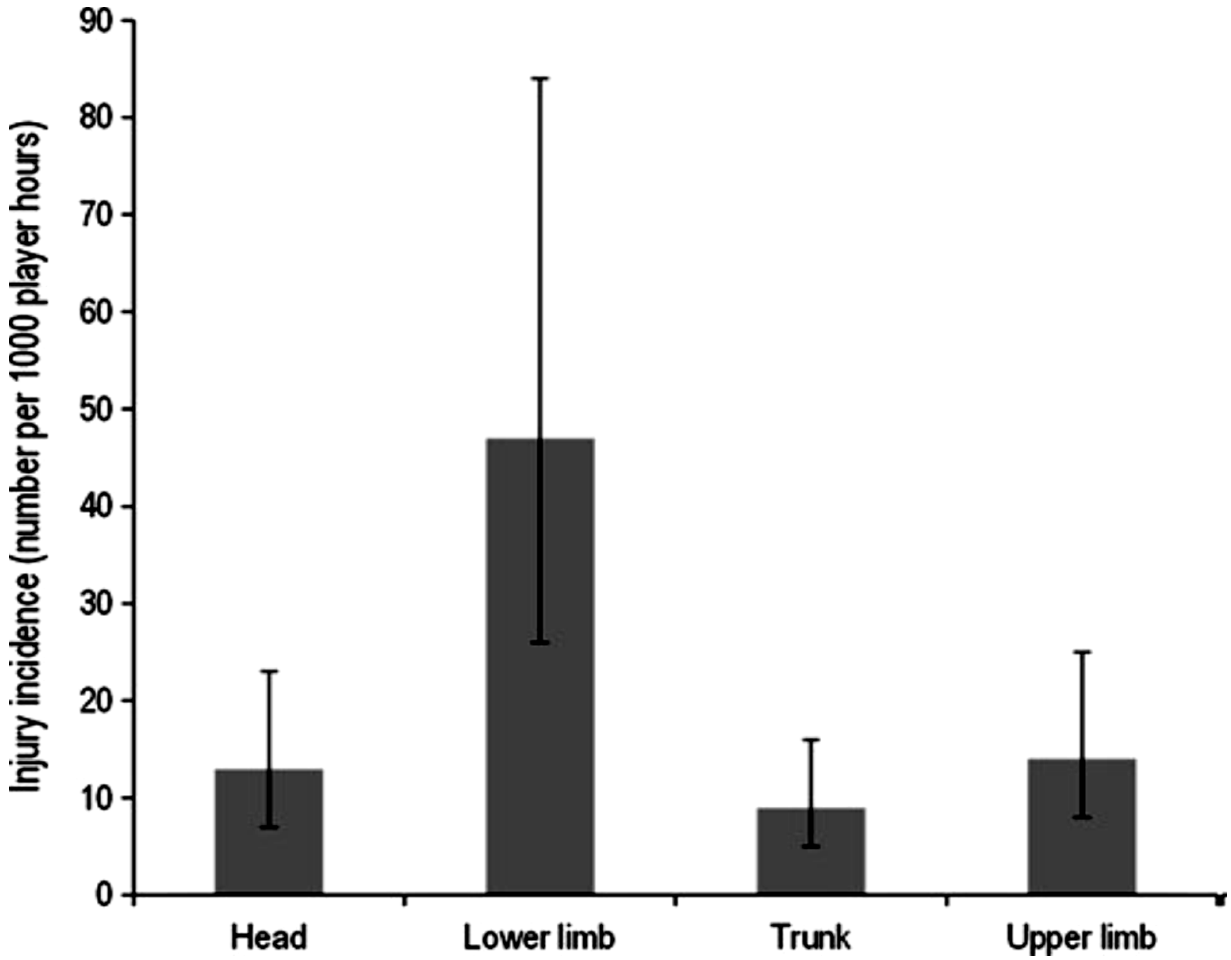


Figure 4. Injury incidence (95 % CI) by injury incident, injury incidence by location of injury. Adapted from Williams et al. (2013).

Injury incidence differs between rugby union playing positions (Figure 6). In a sample of 899 athletes from the top tier of English rugby, hookers had the highest injury rate among the forwards mainly consisting of cervical disk injury but also including significant soft tissue (rotator cuff, meniscal/cartilage, cruciate ligament, calf muscle, hematoma, Achilles tendon and ankle lateral ligament) injuries. Among the backs, centres showed the greatest injury incidence compared to other backs and mainly consisting of hamstring muscle injuries (Brooks & Kemp, 2011). These positional differences in injury incidence further support research into the inclusion of genetic data for the management of player injury.

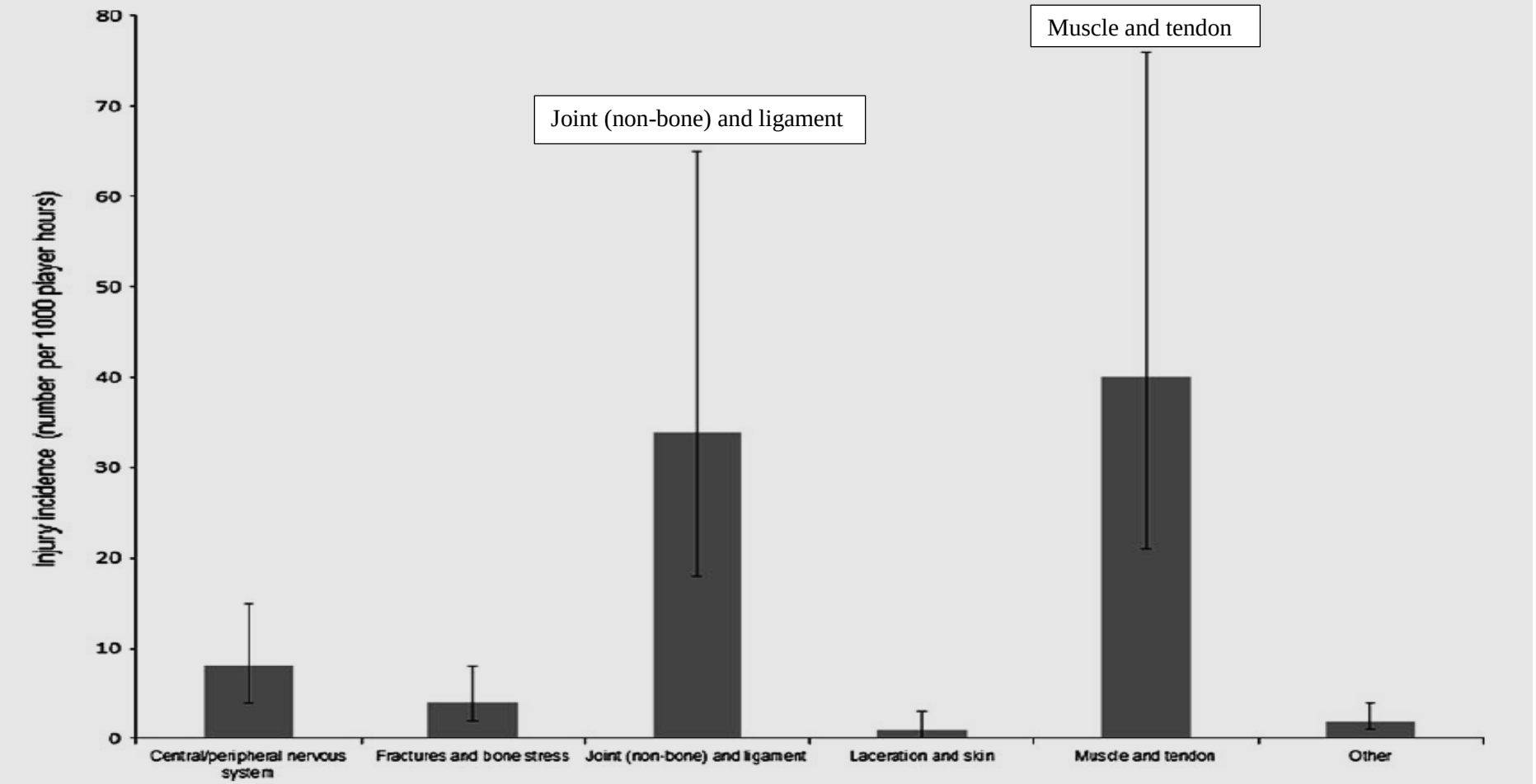


Figure 5 Injury incidence (95 % CI) by injury type. Adapted from Williams et al. (2013).

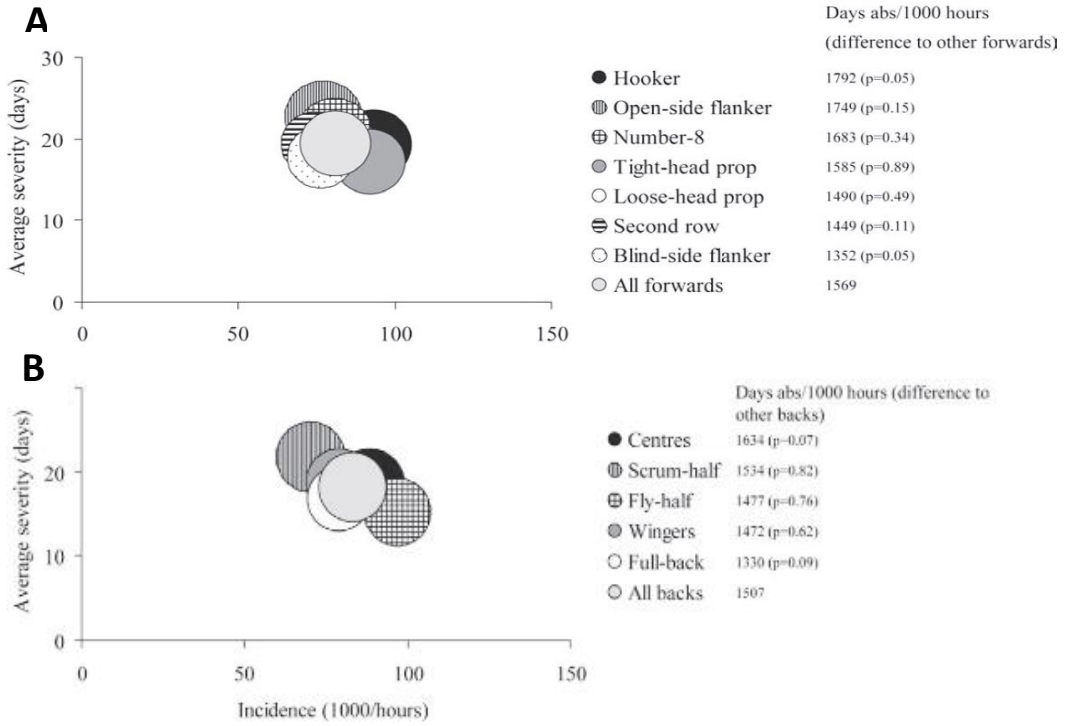


Figure 6 Incidence, average severity and days of absence due to injury for **A**, forwards and **B**, backs. Bubble size: days of absence/1000 player-hours. Adapted from (Brooks & Kemp, 2011)

2.1.7 Concussion (mTBI)

Omalu et al. (2005; 2006), showed for the first time that repeated head injury due to athletic collisions caused significant long term neurological injury which he termed "Chronic Traumatic Encephalopathy" (CTE) and attributed it to repeated mild-traumatic brain injury (mTBI; concussion). As concussion is a form of mTBI and is classified as such (Roozenbeek *et al.*, 2013), concussion will be referred to as mTBI from here on. The long-term effects of mTBI on neurodegenerative illness are unknown (McCrory *et al.*, 2013), however hints towards cognitive decline and depression have been suggested in retired (> 20 years) rugby union players (Decq *et al.*, 2016). These symptoms are similar to those experienced by retired American Football athletes who were subsequently the subject of autopsy, resulting from suicide (~50 years old), in the discovery of CTE (Omalu *et al.*, 2005; Omalu *et al.*, 2006; Omalu *et al.*, 2011). Unfortunately, the only method currently available to identify CTE is post-mortem, leaving these analysis of rugby athletes difficult and unlikely.

Nonetheless, World Rugby have addressed this concern and have implemented strategies to minimise the risk of CTE in rugby athletes (Raftery, 2013; Fuller *et al.*, 2015b; Raftery *et al.*, 2016). Current mTBI incidence for elite 15's rugby union range between 4.6-8.9 per 1000 playing hours, with 40-50% resulting from a tackle situation (Gardner *et al.*, 2014; Cross *et al.*, 2015; Fuller *et al.*, 2015a) - similar to that of most other injuries (Brooks & Kemp, 2011). However, the lower end of this range of mTBI incidence (< 5) is likely to be an underestimation because it is based upon data from several years ago (2012-2013 season), prior to the implementation of the recent mTBI reporting and return to play guidelines (Figure 7). The most recent data from Cross et al. (2015) more truly represents the current rugby mTBI incidence rate and is consistent with the English professional rugby union data from the 2013-2014 and 2014-2015 seasons (Figure 5: England Professional Rugby Injury Surveillance Project Steering Group, 2016). Of particular note was the finding that players diagnosed with mTBI who returned-to-play in the same season had a 60% greater chance of a subsequent time-loss injury than those that did not sustain a mTBI (Cross *et al.*, 2015).

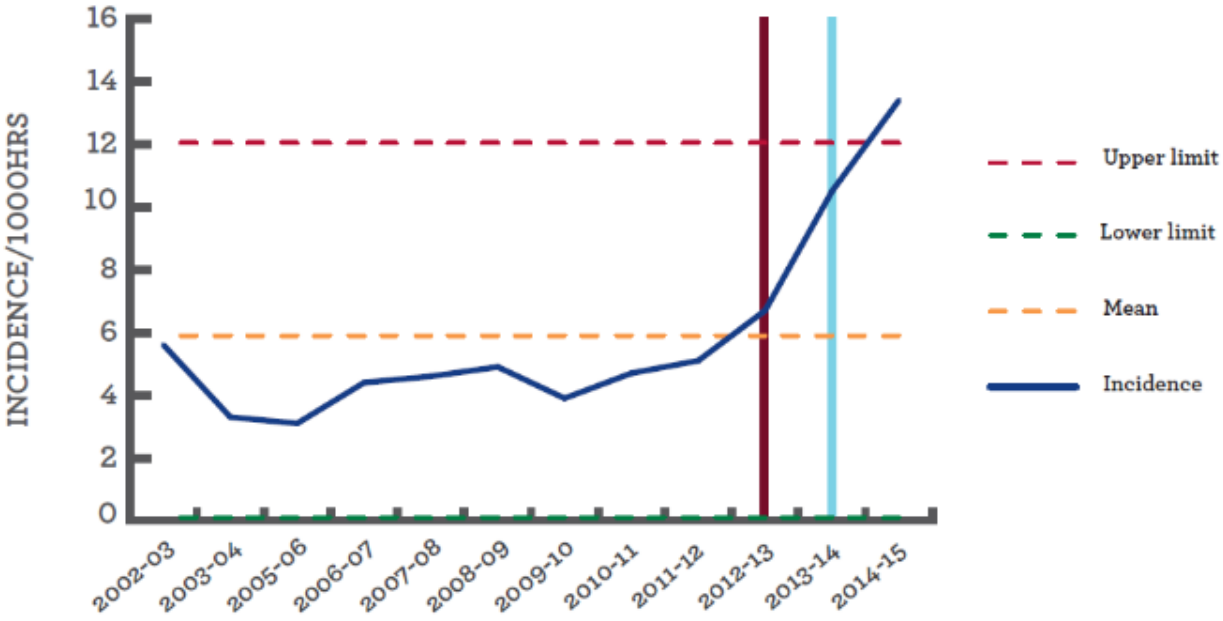


Figure 7 Incidence per 1000 player hours of reported match concussions by season. Vertical blue line denotes change in data collection methods to electronic capture. Vertical **dark red** line denotes introduction of the pitch side concussion assessment. Adapted from the England Professional Rugby Injury Surveillance Project (2016).

Positional variation data in mTBI incidence are scarce and variable. Fuller et al. (2015a) reported no difference between backs and forwards, however the mTBI severity in centres was reported as 105 days absent per 1000 h (Risk ratio = 2.54) and greater than other backs (Brooks & Kemp, 2011). Higher mTBI metrics in centres are supported in Irish international players, with centres showing the second highest number of mTBI (16%) preceded by flankers (backrow; 22%) during the 2010-2011 season (Table 2: Fraas *et al.*, 2014). Again, if there truly is a difference in mTBI rate between playing positions, given the possible catastrophic outcomes (Omalu *et al.*, 2005; Omalu *et al.*, 2006; Omalu *et al.*, 2010; Omalu *et al.*, 2011) and the re-injury rate following a mTBI event (Cross *et al.*, 2015), knowledge of relevant molecular genetic variation among rugby athletes could lead to greater player welfare.

Table 2. mTBI by playing position, Irish Rugby Football Union (IRFU) data. Adapted from Fraas et al., (2014).

Player position	Number (%) of concussions
Prop	9 (9.8%)
Hooker	6 (6.5%)
Lock	12 (13.0%)
Flanker	20 (21.7%)
No. 8	4 (4.3%)
Scrumhalf	11 (12.0%)
Flyhalf	4 (4.3%)
Center	15 (16.3%)
Wing	8 (8.7%)
Fullback	3 (3.3%)
Total	92 (100%)

2.2 Part 2: Sport Genomics

2.2.1 Heritability

Classical genetics is the process of estimating the heritability of a given trait and is investigated by studying families, identical twins (monozygotic; MZ) and fraternal twins (dizygotic; DZ). A number of physiological variables have been investigated in this way and are usually the preceding step to investigating specific target genes (Table 3). For example, Simoneau and Bouchard (1995) showed the heritability of muscle fibre type proportion was ~50% when investigating 58 DZ and 35 MZ twin pairs. Following 20 weeks of endurance training in 98 two-generation families (n = 481), the heritable component of the ability to adapt, specifically an increase VO_{2max}, was ~50% in the well-known HERITAGE Family Study (Bouchard *et al.*, 1999). While these heritability estimates are substantial, other anthropometric phenotypes are estimated to be even greater. For example in a sample of over one million participants, heritability was estimated for body mass index (BMI) and mass at ~60%, height at ~80%, elbow flexion and knee extension strength at ~50% and hand grip strength at ~60% (Silventoinen *et al.*, 2008). Other phenotypes have been given even larger heritability estimates (Table 3; Calvo *et al.*, 2002; Peeters *et al.*, 2007; Missitzi *et al.*, 2008; Busjahn *et al.*, 2009; Schutte *et al.*, 2016).

Table 3 Heritability estimates for selected phenotypes.

Physiological phenotype	Heritability	Reference
Muscle enzymes	~50%	(Bouchard *et al.*, 1986)
% body fat	~30%	(Bouchard & Perusse, 1988)
% fibre type	~50%	(Simoneau & Bouchard, 1995)
Max O_2 uptake	~60%	(Schutte *et al.*, 2016)
Δ Max O_2 uptake	~50%	(Bouchard *et al.*, 1988)
Leg strength	~30%	(Tiainen *et al.*, 2004)
Leg strength	~60%	(Zhai *et al.*, 2004)
Mesomorphy	~80%	(Peeters *et al.*, 2007)
Athlete status	~70%	(De Moor *et al.*, 2007)
Left Ventricular mass	~80%	(Busjahn *et al.*, 2009)
Height	~80%	(Visscher *et al.*, 2006)
Frozen shoulder	~40%	(Hakim *et al.*, 2003)
Tennis elbow	~40%	
Total joint replacement	~50%	(Williams *et al.*, 2015a)
Joint stiffness	~30%	
Knee extension strength	~50%	(Silventoinen *et al.*, 2008)
Hand grip strength	~50%	
Mass	~60%	
BMI	~60%	
Hand grip strength	~60%	(Frederiksen *et al.*, 2002)
Explosive anaerobic power	~70%	(Calvo *et al.*, 2002)
Max power development (5 s)	~70%	
Max O_2 uptake (@2 min)	~90%	
Δ lactate deficit	~70%	
Max EMG activity	~80%	(Missitzi *et al.*, 2008)
Brain plasticity	~70%	(Missitzi *et al.*, 2011)
Muscular degeneration (age-related)	~50%	(Hammond *et al.*, 2002)
Longevity	~20%	(Herskind *et al.*, 1996)
Skeletal muscle mass	~80%	(Livshits *et al.*, 2016)

Max, maximum; BMI, body mass index; EMG, electromyography.

Possibly the most useful future application of predictive genetics within sports genomics will be in the field of injury risk and severity estimation. However, little data exists on the heritability of injuries. Hakim et al. (2003) examined frozen shoulder (FS; characterised by tendon and ligament injury) and tennis elbow (TE; tendon inflammation) in 865 MZ and 963 DZ twin pairs and reported 42% heritability for FS and 40% for TE. More recent research has examined joint stiffness (32%) and fibrotic conditions (connective tissues; 28%) and

even the likelihood of having a total joint replacement (46%; Table 3; Williams *et al.*, 2015a). Additionally, while the author is not aware of heritability estimates for mild-to-moderate traumatic brain injury (TBI), the likelihood of suffering more severe future neurological disorders following these injuries is significant (Goldman *et al.*, 2006). Brain plasticity is a significant factor in the recovery from brain injury (Pearson-Fuhrhop *et al.*, 2009) and has recently been shown to have a heritability of 68% in a small twin study (Missitzi *et al.*, 2011). More specific to the present thesis, De Moor et al. (2007) investigated the genetic component of athlete status in female (1,000 DZ and 793 MZ) twins and reported a heritability estimate of ~70% for athlete status. These data show the considerable genetic component of many phenotypes related to sports performance and investigating this genetic variation is vital to truly understand any physiological trait (Table 3).

2.2.2 Human genetic variation

The variation that exists in the human genome has only relatively recently been documented via progress of the Human Genome Project. Initially, ~11 million single nucleotide polymorphisms (SNPs) and 3 million short insertions and deletions were identified (Sachidanandam *et al.*, 2001; International HapMap Consortium, 2005; Frazer *et al.*, 2007) and, at the time of writing, more than 88 million SNPs have been validated (*http://www.ncbi.nlm.nih.gov/projects/SNP/snp_summary.cgi*). The vast majority of the ~3 billion nucleotides that constitute a human genome do not code for proteins, although most of those nucleotides nonetheless contribute to biological function in some manner via regulation of gene expression (ENCODE Project Consortium, 2012). There are a number of different types of human genomic variation that affect biological function, including rare mutations, structural variations and common SNPs.

The loss of function nucleotide change in the myostatin (*MSTN*) gene is an example of a rare mutation of relevance to human physiological function and results in an absence of the myostatin protein, which is a negative regulator of muscle growth. In humans, only one report of this mutation exists, causing extreme hyper-muscularity and a lean phenotype. At six days old, the affected child, presenting as homozygous for the mutation, showed twice the quadriceps muscularity (cross-sectional area) than that of an age-matched control (6.7 cm^2 compared to 3.1 cm^2; Schuelke et al., 2004). The negative impacts (if any) of this mutation have not yet been reported and at 4.5 years the child showed no pathological symptoms. However, because myostatin affects other muscle types including cardiac muscle, later-onset pathological symptoms resulting from excessive myocardial growth are a possibility.

Genetic variations where an allele occurs relatively commonly (≥ 1%) are called polymorphisms, as opposed to mutations. An example of one such polymorphism relevant to exercise physiology is the insertion/deletion (I/D) variation on intron 16 of the angiotensin converting enzyme (*ACE*) gene on chromosome 17. The *ACE* gene encodes the angiotensin converting enzyme, which is the main active product of the rennin-angiotensin system. The insertion of the 287 base pair (bp) DNA sequence (I allele), despite being in a non-protein coding region of the *ACE* gene, produces lower circulating and tissue concentrations of the ACE protein compared to the D allele (Almeida *et al.*, 2010). One main action of the ACE enzyme is to degrade inactive angiotensin I and generate the vasoconstrictor angiotensin II, while another is to degrade vasodilator kinins and is therefore hypothesised to influence mitochondrial oxygen consumption and exercise economy in some circumstances (Puthucheary *et al.*, 2011).

An example of a SNP relevant to exercise performance found in an exon (a region of DNA that encodes for protein) is the *ACTN3* R577X gene variant. The *ACTN3* gene encodes for the α-actinin-3 protein, which is expressed almost exclusively in fast glycolytic type II fibres and is a structural component that binds the actin thin filament to the sarcomere Z line. This SNP is located on exon 16 of the *ACTN3* gene (North *et al.*, 1999) and is characterised by the replacement of the normal codon (a 3 bp sequence that codes for an amino acid (Arg; R)) by a premature termination codon (X) at the 577 amino acid position and results in the complete absence of the α-actinin-3 protein. The absence of this protein (XX genotype) is associated with a lower proportion of type II muscle fibres - amongst other fibre characteristic - and, accordingly, is found at a lower frequency in elite power/sprint athletes compared to other athletes and non-athletes (Eynon *et al.*, 2013a). Both the *ACE* I/D and *ACTN3* R 577X variants will be discussed in greater detail in *part 3* of this chapter as they are the subject of chapter *4*.

2.2.3 Genetics and athlete status

In a review published in 2012, at least 79 genetic markers were identified that had been associated, in at least one prior research paper, with elite athlete status (Ahmetov & Fedotovskaya, 2012; Ahmetov *et al.*, 2016). That number was reduced to 20 when the criterion was at least two prior research papers, and probably even some of those associations will not prove to be true as more data are accumulated. The first scientific investigation to assess the molecular genetic component of elite athlete status (Gayagay *et al.*, 1998) showed a significant association of the *ACE* I/D variant (mentioned previously) with elite status in 64 rowers. Subsequently, the *ACE* I allele has been associated with elite performance in a variety of sports, though not consistently, and the research is probably best summarised by reference to the meta-analysis of Ma et al. (2013) who found that the II genotype of *ACE* I/D

was associated with physical performance (Odds ratio (OR) = 1.23), especially endurance performance (OR = 1.35).

For the *ACTN3* R577X variant already mentioned, the R allele has been consistently associated with elite power and sprint athletes from a variety of backgrounds and in top elite sprinters, a complete absence of the XX genotype has been identified (Yang *et al.*, 2003; Niemi & Majamaa, 2005), compared to an XX genotype frequency of ~18% in the general Caucasian population. Again, the meta-analysis by Ma et al. (2013) nicely summarises the association of the R allele with elite power athlete status (OR = 1.21).

A number of athlete cohorts have gradually emerged, hosting steadily larger samples of elite and sub-elite athletes from various sports for the investigation of athlete status (Pitsiladis *et al.*, 2013). A UK athlete cohort is currently being established, of which the samples within this thesis are part (RugbyGene project; Chapter *3.1.1*). Furthermore, an advancement of sports genomics has seen the inclusion of team sport athlete data, however considerable methodological limitations have been identified and are discussed in the following.

2.2.3.1 Genetic and team sports.

Unlike individual or single attribute based sports (e.g. sprint speed in sprinters, max oxygen uptake in endurance athletes), the evidence for genetics playing a role in elite status within team sports is less compelling (for a full overview of each sport see appendix 5). Consistent with individual sporting performance, the most convincing evidence within team sport genomic investigations are of *ACTN3* gene polymorphisms. Furthermore, evidence form soccer athletes shows the most compelling genetic associations between elite team sport athlete status and *ACTN3* (Santiago *et al.*, 2008; Egorova *et al.*, 2014). Other genetic associations are certainly less convincing (see limitations below), for example low participant numbers and to a lesser extent the classification of “elite” is questionable in a

number of studies that have investigated associations of *ACE* I/D, *TNF* G308A and *FGFR1* genes with team sport athlete status (Gronek *et al.*, 2013; Salles *et al.*, 2015). Similar to the coverage of genetics of elite status in these aforementioned team sports, rugby is also poorly described with little convincing, yet potential, evidence for the specific role played by genetics and is discussed below.

2.2.3.2 Genetics and rugby

As early as 1922 scientists were hypothesising about the role of heritability in rugby union athletes. Jack (1922) documented the playing positions in 23 sets of elite rugby-playing brothers (n = 63), including a number of international representatives and concluded that *"the ability required for playing in certain positions in rugby football is inherited*" (page 161). Five full publications in peer-reviewed journals applying molecular genetics to rugby union exist (appendix 5). Goh et al. (2009) reported that the *ACE* II genotype was associated with a higher ventilatory threshold in non-elite Asian rugby players but the very small cohort (n = 17) is a major limitation to that study. Bell et al. recently assessed *ACE* I/D and *ACTN3* R577X genotypes in 68 and 102 young non-elite rugby union players, respectively (Bell *et al.*, 2009; Bell *et al.*, 2010; Bell *et al.*, 2012c). No associations were identified between either polymorphism and athlete status, playing position or the physiological and anthropometric parameters assessed, perhaps due to the rather small sample size and the sub-elite status of the players. However, a statistical tendency was identified within playing positions for *ACTN3* (P = 0.066; Bell *et al.*, 2012c), which suggests that with a greater sample size and higher standard of player statistical significance may be reached. With such little data regarding the genetic characteristics of rugby union players in the literature, this gap should be filled with high quality data using appropriate experimental designs (as per section *2.2.6*).

2.2.4 Limitations in team sport genetics

The scientific literature has recently seen a substantial rise in the number of peer reviewed sports genomics publications (Ahmetov & Fedotovskaya, 2012; Ahmetov *et al.*, 2016). Because of this increase in publication volume, multiple reviews have been published concerned with updating the readers on the advancements in the field (Rankinen *et al.*, 2001; Pitsiladis *et al.*, 2013; Bouchard, 2015; Loos *et al.*, 2015). To date, a large portion of reviews have focused specifically on the genetic variation of either endurance athletes (Eynon *et al.*, 2011c; Wilber & Pitsiladis, 2012), strength/power/speed athletes (Berman & North, 2010; Hughes *et al.*, 2011; Eynon *et al.*, 2013a) or a combination of the two (Puthucheary *et al.*, 2011; Ma *et al.*, 2013; Tucker *et al.*, 2013). However, there has been little synthesis of sports genomic data in relation to team sports such as soccer, rugby, field and ice hockey, – i.e. those that do not also exist as individual sports such as relay swimming or running, doubles badminton or tennis, etc.

Additionally, recent publications have exposed inconsistencies in the reporting of important statistical analyses within the peer-reviewed literature of well-known academic journals such as Nature Genetics, American Journal of Human Genetics, American Journal of Medical Genetics, Cell and others (Salanti *et al.*, 2005; Kirkham & Weaver, 2015; Namipashaki *et al.*, 2015; Chavalarias *et al.*, 2016). These inadequacies have been suggested to result in false findings despite appearing as statistically significant results (Ioannidis, 2005). An investigation of this kind has never been performed in the context of sports genomics and could identify important, yet rectifiable errors within the scientific literature.

It is important for sport and exercise genomic research to be of the highest possible standard (better than its current state; Pitsiladis *et al.*, 2013) and to learn from the mistakes of other fields of genetic investigation (Mattsson *et al.*, 2016). These limitations might have substantial implications for ethical matters (Wackerhage *et al.*, 2009) such as familial

sensitivity in relation to personal genetic data and magnified by the possibility of misunderstanding of particular health-associated genomic information (Webborn *et al.*, 2015). Improving the quality of team sport genomic research can be achieved by reviewing the currently available literature, identifying persistent methodological limitations that can be rectified and uniformly adopted/applied across the field. Therefore, the purpose of this section was to review team sport genomic publications (appendix 5) and to identify methodological concerns specific to the genomic investigation of team sport athletes and attempt to rectify these limitations in the experimental chapters of the present thesis.

2.2.4.1 Summary of findings

The large number of sports genomic publications that were identified can be categorised into those that include team sport athletes (Multi Sport Articles), normally pooled into a single group (mixed) for comparison with either endurance or sprint/power/strength athletes, and those that solely investigate individually team sport athletes (Single Sport Articles; see appendix 5). Of these team sport publications, a total of 67 original articles were identified with 27 assessing only individual team sports within the analysis, the majority of those being soccer athletes (55.5%) followed by rugby union (18.5%), with the rest divided between four other sports (Figure 8). These data show that of the limited scientific attention devoted to sports genomics, the majority of these involve soccer athletes - with little investigation of other team spots athletes. It is clear that more scientific efforts need to be devoted to the genomic variation of other team sport athletes and the present thesis attempts to do this regarding elite rugby union athletes.

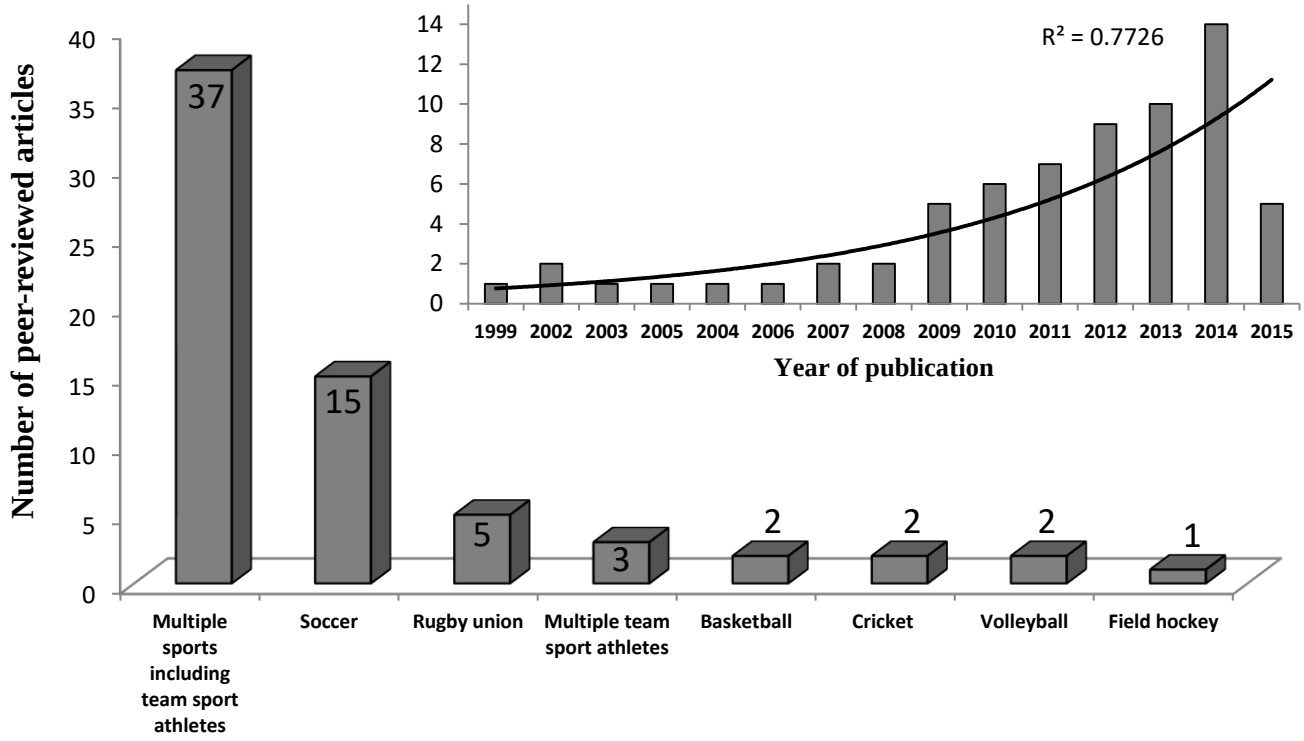


Figure 8 Number of sports genomics publications until December 31st 2015 that include team sport athletes, categorised by sport. **Insert**: Number of sports genomic publications that include team sports by year. Regression coefficient describes an exponential curve (*Microsoft Excel, 2013*).

Figure 8 (insert) further documents the yearly growth (R^2 = 0.77) in team sport genomic publications, following 2014, with the greatest growth occurring from 2011-2014 and a reduction in 2015. This drop in publications following 2014 is possibly a reflection of the pattern change in sport genomic research driving towards research collaboration and generation of larger athlete cohorts (Pitsiladis *et al.*, 2016). Furthermore, it is noteworthy that of the single sport articles identified, 51.9% of the publications had a sample size ≤ 100, 33.3% had ≤ 201 and 14.8% showed > 200 (Table 4). This identifies a significant issue in the field of sports genomics, however attempts are being made to combat this issue with largescale internationally collaborative consortia (Pitsiladis *et al.*, 2016).

In relation to statistical analysis within the published team sport genomic research, considerable shortcomings were identified (Table 4). Hardy-Weinberg equilibrium was absent in ~36% of all reviewed articles and in 37% of individual team sports publications. Reporting some form of effect estimate was present in only 29% of all reviewed articles and 34% of single team sport genomic original investigations. However, the most concerning result pertains to the extremely low consideration of multiple testing errors, with only 31% and 33% of articles reporting corrections (Table 4). Thus ~70% of peer-reviewed sports genomic investigations containing team sport athletes may be fundamentally flawed (assuming that ≥ 2 null-hypothesis are performed) and contain type 1 errors. Given these methodological limitations and that team sport specific genetic publications are generally growing (Figure 6 insert), future research within sports genomics needs to improve on quality and correct these simple inadequacies.

Table 4 Summary of statistical and sample size findings for all reviewed articles and individual single sport analysis articles.

	Statistical tests	**All articles n = 67 (%)**	**Single sport analysis n = 27 (%)**
Hardy-Weinberg	Chi Square	43 (64.2)	17 (63.0)
Effect size	Regression analysis, likelihood ratio and Odds ratio (OR)	22 (32.8)	9 (29.6)
	OR specifically	19 (86.4)	7 (77.8)
Multiple correction	FDR, Bonferroni, Holm–Bonferroni, Benjamini-Hochberg and Scheffe's multiple comparison	14 (21.9)	5 (19.2)
	FDR specifically	3 (4.7)	2 (7.7)
Athlete sample size	**n ≤ 100**	21 (31.4)	14 (51.9)
	n = 101 - 200	24 (35.8)	9 (33.3)
	n ≥ 201	22 (32.8)	4 (14.8)

These data are generated under the assumption that all reviewed investigations required the above statistics analysis. Multiple corrections data includes investigations undertaking ≥ 2 null-hypothesis test for the genetic analysis, i.e. for all articles n = 64 and for single sports analysis n = 26.

2.2.5 Considerations based on team sport review findings

2.2.5.1 Sex

Striving to identify genetic variants important for athlete populations have led researchers to combine male and female athletes in their genetic analysis (appendix 5). However, because athleticism is highly phenotypically specific, consideration of the physiological diversity between male and female elite athletes must be accounted for. As such, while genetic variation may be similar between sexes, autosomally speaking, physiological demands and capabilities differ considerably at the respective elite levels (Reilly & Borrie, 1992; Stølen *et al.*, 2005; Gabbett, 2007; Marques *et al.*, 2009; Ziv & Lidor, 2009; Lidor & Ziv, 2010; Ziv & Lidor, 2010; Lidor & Ziv, 2015; Morehen *et al.*, 2015). Because of these physiological differences, combining male and female elite athletes is not appropriate and may introduce false negatives into study results. If, when both sexes are assessed individually in genotype-phenotype and case-control associations involving the same variant(s), combining the sexes may become viable and warranted. Until then, assessing both male and female elite athletes separately (Holdys *et al.*, 2013) and for a given individual team sport or athletic category (sprint/power or endurance) is the only way to reduce the possibility of diluting the experimental findings due to physiological diversity between elite athletes of different sexes. Nonetheless, combining males and females in the control group is appropriate, as long as the frequency of the genetic variants being studied does not differ between sexes, with the genetic similarities detailed within the scientific report.

2.2.5.2 Eliteness

Classifying athletic level is a considerable challenge regarding team sport athletes and a sport specific standardised method that could be easily replicated - therefore easily comparable between studies - would give the field greater transparency regarding the results

from individual investigations. Current findings regarding the genetic contribution to athletic status have so far been generally inconclusive (Pitsiladis *et al.*, 2013) and might, partly, be a result of the variation in athletic level and the combination of athletic levels within a given cohort/group/category. Thus, there is a necessity to define the highest athletic level (eliteness) for individual sports within a given scientific investigation. For example, are soccer athletes competing in the top Spanish league (Pruna *et al.*, 2013) equivalent to Polish top league athletes (Egorova *et al.*, 2014)? As such, is it possible to compare their genetic association results? Within rugby union the author proposes a definition of 'elite' as athletes competing in the highest competitive league of a 'Tier 1' rugby nation (detailed in chapter *3.1.1*), originally defined by World Rugby (the governing body for rugby union; International Rugby Board, 2004). While it is appreciated that defining 'elite' in this way may be difficult for some sports, defining eliteness for a given sport and all following research adhering to that perimeter would significantly enhance the confidence in scientific findings and allow genuine replication.

2.2.5.3 Sample size

The nature of genetics research dictates that extremely large sample sizes are needed before genuine conclusions can be made about the variation within a given cohort, for a given phenotype (Gauderman, 2002). Due to the natural rarity of elite athletes it is difficult to congregate large samples of particular athlete groups (power/endurance/team sport athletes) and even more difficult to recruit large cohorts of single sport athletes (such as Rugby Union). As such, intentions to recruit large cohorts have led some authors to combine 'a single athlete' from a number of sports (all with considerable physiological differences, for example shooting (n = 1), cycling (n = 1), jazz dance (n = 1) etc.) within the cohort for genetic analysis (Boraita *et al.*, 2010; Jin *et al.*, 2015) and invariably, have led to misleading

findings. This type of athlete inclusion is analogous to taking 1000 ‘sick’ people (suffering from various medical conditions) and looking for the ‘sick gene’.

Nevertheless, to overcome this issue of sample size, a number of athlete cohorts have emerged hosting large cohorts of elite, sub-elite and non-elite athletes from various sports (Pitsiladis *et al.,* 2013). Some of these cohorts combine to reduce the likelihood of type 1 error. The largest of these merged cohort studies included 2178 Caucasian athletes of Russian (n = 1780) and polish (n = 398) ancestry, of which 31% were classified as elite (Mustafina *et al.,* 2014). While this design rightfully attempts to combat the issue of sample size, it may also increase the likelihood of type 1 error. In fact, estimations for adequate sample size have been postulated for both single SNP case control and genome wide studies (Hong & Park, 2012). Depending on the genetic model of interest, the minor allele frequency and statistical power of 80%, a sample of 248 cases (team sport athletes in this case) and the same in controls are the minimum requirements. Acquiring this number of elite samples is difficult enough, however in the case of team sports athletes where the physiological demands differ considerably across positional groups, this number is required for each individual group (playing position) in order for the results to be conclusive. Due to the difficulties in recruiting large cohorts of homogenous athletes (i.e. the same team sport), preliminary investigations should include approximately this number (n = 248) of elite cases and controls for initial analysis – which is often not the case (Table 4). Furthermore, the development of the cohort should eventually include this sample size within each positional group. For example, basketball comprises three positional groups (centres, guards and forwards) and therefore would ultimately require ~744 cases to test a single polymorphism among playing positions. This number of participants would give the analysis the appropriate statistical power for one to be confident in the research findings - that the studied polymorphism was truly associated (or not) with basketball athlete status. However

acquiring a sample of this size of elite (often professional) team sport athletes is extremely difficult and would require many years to accomplish.

2.2.5.4 Geographical ancestry considerations

Geographical ancestry are important considerations for case-control and genotype-phenotype association studies (Jorde & Wooding, 2004) because individual genetic variants are known to differ in allele frequency between populations (1000 Genomes Consortium, 2012). Therefore, investigations of molecular genetic markers should be performed on athletes from well-defined geographic ancestries. Only when one understands the genetic diversity of individual geographical ancestries can one compare ethnic diversities for a given variant. Current team sport genetic publications, for the most part (however not exclusively, for example Pruna *et al.*, 2013), consider geographical ancestry in their study design, but is still a substantial consideration worth mentioning within the context of the present thesis.

2.2.5.5 Positional analysis

Investigating the genetic components of team sport athleticism provides the unique consideration of positional differences, in terms of physiological demands. These differences can be quantified by game demand data (for rugby union see chapter 2.1.2; Table 1) that shows a preference for differing metabolic pathway proportions dependent on playing position (Bradley *et al.*, 2013; Jones *et al.*, 2015). Furthermore many physiological quantities (maximal strength, running speed, aerobic capacity, muscle power, etc. for rugby union see Table 1) differ by playing position (Abdelkrim *et al.*, 2010; Smart *et al.*, 2013) and should be considered as these quantities may have considerable genetic components. Three recent investigations of elite athletes have considered positional specificity in their genetic analysis of team sport athletes (Ginevičienė *et al.*, 2009; Egorova *et al.*, 2014; Gineviciene *et al.*, 2014). When soccer athletes were arranged into positional groups (forwards, defenders,

midfielders and goalkeepers), Gineviciene et al. (2014) identified genetic variation across playing position for *ACE, PPARGC1A* and *PPARA*. Similar positional variation with *ACE, ACTN3, PPARA, UCP2* was associated in Russian soccer athletes (Egorova *et al.*, 2014). However, an earlier investigation showed no difference in terms of *ACE* I/D variation (Ginevičienė *et al.*, 2009). These data show the importance of positional analysis when considering team sport athletes - where the physiological and positional demands differ (further evidence is presented in chapter *4* and *5*).

2.2.6 Statistical aspects

Molecular genetic association studies aim to investigate specific genetic variations in a population, which have a phenotypic consequence. Applying the correct statistical analyses and interpretation of the results generated from these analyses is paramount to allow meaningful conclusions to be drawn from the research findings. However, lacklustre use and reporting of statistical analysis are a considerable concern across the biomedical literature (Table 4; Salanti *et al.*, 2005; Kirkham & Weaver, 2015; Namipashaki *et al.*, 2015; Chavalarias *et al.*, 2016). Nevertheless, a number of small but important statistical considerations can strengthen the confidence of athlete-genomic research findings and enhance the progress of the field.

2.2.6.1 Hardy-Weinberg principle

The Hardy-Weinberg principle (HWP; Hardy, 1908; Weinberg, 1908) states that in a randomly mating population, genotypic frequencies are indicative of the sample function of allele frequencies and in the absence of perturbing forces, such as selection, genetic drift, mutation and migration will remain constant over time (Waples, 2015). Because mutations are extremely rare and genetic drift is a function of sample populations (occurs in small isolated populations; Waples, 2015) and team sport genetic samples come from the wider

population, these indicatives can be ignored in the context athlete genomics. Regardless, a sample that deviates from this principle is not indicative of the wider population and is likely due to population stratification (Wigginton *et al.*, 2005) or genotyping error (Xu *et al.*, 2002). It is important to assess the HWP to insure that the population of interest is representative of the general population, specifically within the control sample. In fact, Ziegler et al. (2011) recommended that in case-control study designs the HWP should be assessed in the control group only because a deviation from HWP in 'cases' may indicate a genetic association. Considering athlete-genome assessment, researchers are artificially selecting cases (athletes) aiming to identify the presence or absence of local population stratification. Testing these groups, to be an indicative sample of the general population, is rather meaningless in a case-control design (Ziegler *et al.*, 2011). However, when investigation genotype-phenotype associations (in the absence of a control group) or in case-control athlete group(s), testing for the HWP should be applied as an indication of genotyping error, as population stratification is unlikely because most sport genomic investigations control for geographical ancestry. In the present chapter of team sport genetic association studies (n = 67), 36% did not report adherence to or deviation from HWP which is similar to that reported in other fields of genetic investigation - in highly prestigious journals (Salanti *et al.*, 2005; Namipashaki *et al.*, 2015). It is important that the field of sports genomics overcomes this simple oversight, which does appear to be the case in the more recent publications (appendix 5).

2.2.6.2 Corrections for multiple hypothesis testing

Following analysis of the HWP and the appropriate null-hypothesis test, the consideration of type 1 error (discovering false positives), through multiple testing corrections, is often neglected in sports genomic investigations (Table 4). Given the large number of statistical tests often applied in genomic investigation, controlling for false positives is of paramount

importance to generate valid results. Essentially, the greater the number of statistical tests performed, the greater the likelihood of discovering false positives leading to spurious and often un-reproducible results. There are two main ways of controlling for this error, the *family-wise error rate* (FWER; such as the Bonferroni method, 1936) or the *false discovery rate* (FDR; such as Benjamini & Hochberg method; Benjamini & Hochberg, 1995). Often, the FWER, which effectively makes the confidence coefficient 99% (rather than the traditional 95%) may be too strict, as the amount of null-hypothesis tests increases (Noble, 2009). This, may lead to the occurrence of type 2 errors (failing to detect associations) and as such the FDR is the preferred method (Noble, 2009). The FDR takes into account the statistical significance of the tested null hypotheses as a ranked proportion of the number of tests performed (for a more comprehensive view of multiple testing procedures, see Dudoit & Van Der Laan, 2007) and in doing so also controls FWER (Benjamini & Yekutieli, 2001). There have been multiple variations of FDR tests developed (reviewed in Austin *et al.*, 2014) since the original proposal (Benjamini & Hochberg, 1995). Perhaps the most interesting, and applicable to the present chapter, is the recently proposed hierarchical testing procedure which allows for the control of FDR and provides a more reliable basis for the identification of variant associations while maintaining statistical power (Peterson *et al.*, 2016). The Peterson et al. method allows for the hypotheses tested to be subgrouped into "*families*" of hypotheses to investigate specific scientific questions which may relate to the specific phenotype being tested. This method is worthy of consideration for future reports, specifically within team sport investigations when positional considerations will require additional statistical tests. Presently, only 22% of team sport genetic studies reported any form of multiple testing corrections (Table 4). This is considerably lower than other biomedical fields of investigation (72%; Kirkham & Weaver, 2015). Furthermore, only three studies (4.7%) reviewed within the present chapter considered FDR (Table 4). Given the importance of identifying valid results, controlling for type 1 and 2 errors is of principal

importance for sports genomics investigation and the best method of doing this currently stands with FDR testing.

2.2.6.3 Effect size

It has long been suggested that the null-hypothesis testing approach (such as the P value) is an inappropriate statistical approach for drawing conclusions from scientific investigations, at least as an exclusively reported value. One journal 'Basic and Applied Social Psychology' have banned the reporting of P values from any future submissions and requires every submitted article to provide some measure of effect size (Trafimow & Marks, 2015). An effect size can be defined as *"a quantitative reflection of the magnitude of some phenomenon that is used for the purpose of addressing a question of interest"* (page 137; Kelley & Preacher, 2012). This can be interpreted as the extent (magnitude) of an association that a genetic variant(s) has on a given parameter (athletic status, positional specificity, phenotypic measures) within a team sport, and can be measured with an effect statistic. There are a number of ways to estimate the effect size of an association, for example relative risk, odds ratio (OR) etc. risk ratios for example, can only be calculated when groups intended for comparison are from the same population sample (as in, the effect of drug A or B in a disease). Whereas in a case-control design when one of the cohorts has been specifically selected (athletes) and is being compared to a completely separate cohort (control), the appropriate effect static is the OR (Clarke *et al.*, 2011), with equivalent effect estimates considered when genotype-phenotype study designs are applied (Lakens, 2013).

Within the current sample of team sport genomic inquiries, only 33% of the reviewed articles presented findings pertaining to some form of effect size with 86% of those 33% reporting OR (Table 4). Interpretations of OR statistics have been recently investigated, with the magnitude of the observed effect suggested as small (OR ≤ 1.68), medium (OR ≤ 3.47) and

large (OR ≤ 6.71; Chen *et al.*, 2010). Furthermore, an OR < 2.2 has been identified as having low predictive efficiency in terms of binary genetic analysis and an OR > 5.4 presenting as high (Rubanovich & Khromov-Borisov, 2014). If a large OR is identified with appropriately tight confidence intervals this will be an indication that the sample size was effective and the effect was real. As such, within sports genomic investigations it is no longer acceptable to report significance values without some measure of the effect. Furthermore, the interpretation and discussion of OR results should be centred on greater predictive ability of OR statistics (i.e. OR ≥ 5.4).

2.3 Part 3: Rationale for thesis candidate gene variants

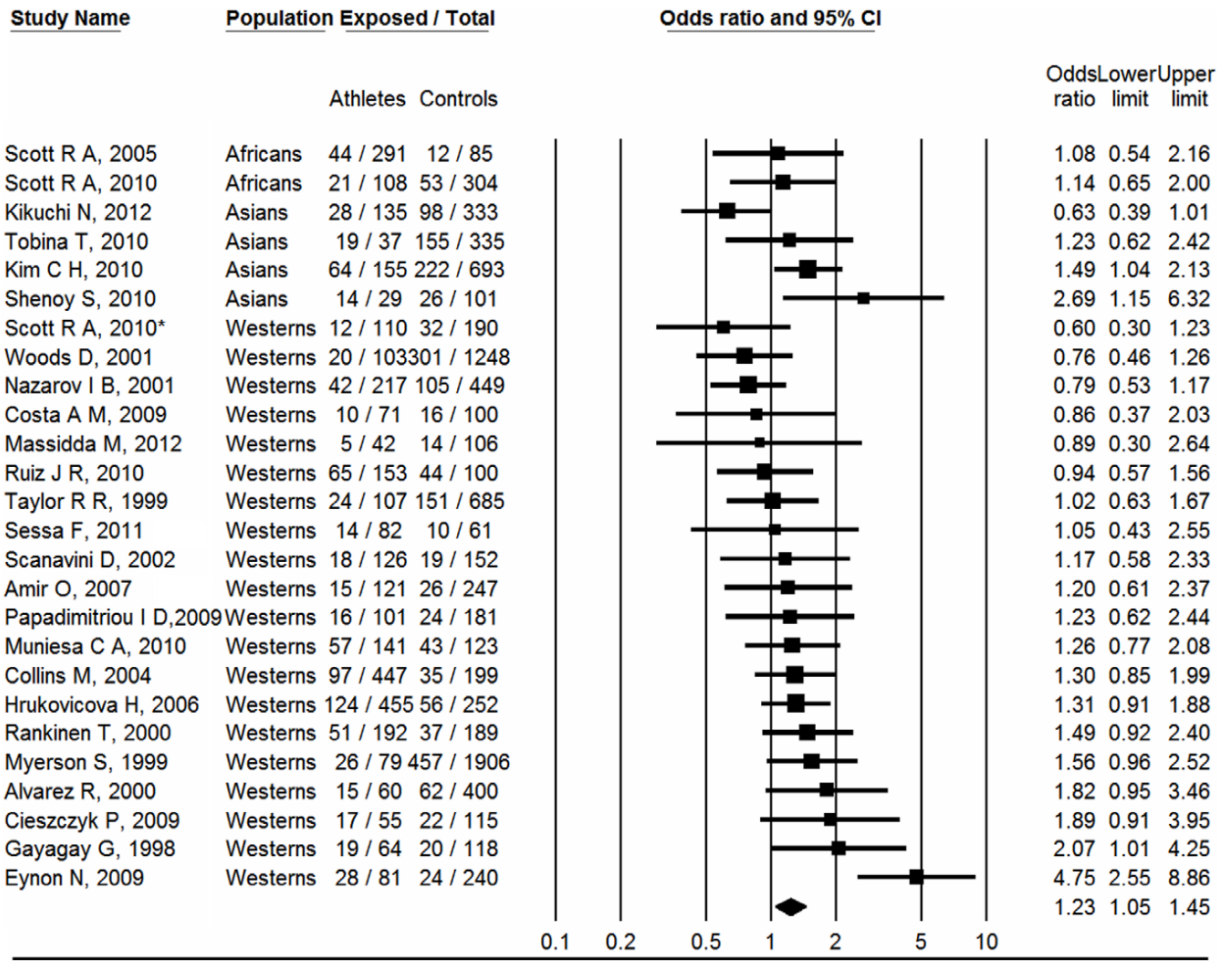
2.3.1 *ACE* I/D and *ACE* tag SNP rs4341

Angiotensin converting enzyme (ACE) degrades the inactive decapeptide angiotensin I and generates the vasoconstrictor octapeptide angiotensin II (Dzau, 1988a; Munzenmaier & Greene, 1996), while accelerating the degradation of vasodilating kinins (Dietze & Henriksen, 2008). The ACE enzyme is the main active product of the rennin-angiotensin system (RAS; Erdös & Skidgel, 1987), which is the system responsible for control and regulation of blood pressure/volume and exists in a number of tissues. Local RAS has been identified in adipose tissue (Jonsson *et al.*, 1994), human myocardium (Dzau, 1988b), and skeletal muscle (Reneland & Lithell, 1994). Interestingly, in skeletal muscle, angiotensin II has been shown to modulate muscle hypertrophy in response to mechanical load (Gordon *et al.*, 2001) and appears to regulate smooth (Geisterfer *et al.*, 1988; Berk *et al.*, 1989) and cardiac (Sadoshima *et al.*, 1993; Ishigai *et al.*, 1997) muscle growth.

A common *ACE* gene variant has been identified and is characterized by the presence (insertion; I allele) or the absence (deletion; D allele) of a 287-bp sequence in intron 16 of chromosome 17, representing an Alu repeat element. Furthermore, an *ACE* tag SNP (rs4341) has been identified in perfect linkage disequilibrium with *ACE* I/D in Caucasians (Glenn *et al.*, 2009) and Asian (Tanaka *et al.*, 2003) populations and is now commonly used. The I allele has been associated with lower circulating (Rigat *et al.*, 1990; Almeida *et al.*, 2010) and myocardial tissue (Danser *et al.*, 1995) ACE activity and a higher proportion of slow-twitch type I skeletal muscle fibres (Zhang *et al.*, 2003). In humans, *ACE* genotype has been associated with cardiac and skeletal muscle hypertrophy in response to exercise training (Montgomery *et al.*, 1997; Folland *et al.*, 2000). Specifically, the D-allele has been repeatedly associated with increased left ventricular mass following training in military

recruits (Montgomery *et al.*, 1997), endurance athletes (Di Mauro *et al.*, 2010) and elite footballers (Fatini *et al.*, 2000).

In the context of human physical performance, the *ACE* I allele has been associated with elite endurance performance in a variety of events (Gayagay *et al.*, 1998; Montgomery *et al.*, 1998; Myerson *et al.*, 1999; Alvarez *et al.*, 2000; Cieszczyk *et al.*, 2009; Cieszczyk *et al.*, 2010), with the D allele being associated with sprint and power-related sports (Woods *et al.*, 2001; Kikuchi *et al.*, 2012; Eider *et al.*, 2013; Papadimitriou *et al.*, 2016). However, several other studies have reported no association between *ACE* I/D and athlete status, while in Israeli athletes the D allele is overrepresented in endurance athletes compared to sprinters (Amir *et al.*, 2007), possibly due to different geographic ancestry, accidental selection bias or other *ACE*-related molecular interactions (Raleigh, 2012). In terms of association studies of mixed metabolically demanding sports, the D allele appears more prevalent, but only in soccer athletes (Juffer *et al.*, 2009; Egorova *et al.*, 2014; Gineviciene *et al.*, 2014), with no difference in others (Bell *et al.*, 2009; Bell *et al.*, 2010). Furthermore, in a recent meta-analysis, Ma et al (2013) reported the II genotype was associated with physical performance (OR = 1.23), especially endurance performance (OR = 1.35; Figure 9), justifying the inclusion of *ACE* I/D in the present thesis (Table 5). Interestingly, Gineviciene et al. (2014) showed that *ACE* DD genotype frequency was lower in defenders (P = 0.033) and midfielders (P = 0.012) compared to controls, suggesting the existence of positional variation. Given that *ACE* D allele appears more prevalent in elite soccer athletes, it is plausible that this relationship may extend to rugby athletes and warrants investigation (see experimental chapter *4*).



Subgroup analysis		Summarized OR (95% CI)	No. of included studies	Heterogeneity test	
				I^2(%)	*p*
All		1.23 (1.05-1.45)	26	51.54	<0.01
Gender	Male	1.10 (0.90-1.35)	10	19.79	0.26
	Female	0.69 (0.37-1.26)	2	0.00	0.63
Ethnicity	Westerns	1.25 (1.04-1.50)	20	51.86	<0.01
	Asians	1.26 (0.72-2.18)	4	74.86	<0.01
	Africans	1.12 (0.72-1.73)	2	0.00	0.91
Sport Discipline	Endurance	1.35 (1.17-1.55)	17	0.00	0.66
	Power	0.93 (0.64-1.34)	13	75.19	<0.01

Figure 9 Meta-analysis of the association between sport performance and *ACE* I/D polymorphism (II vs. ID+DD). Abbreviation: CI, confidence interval; OR, odds ratio. Adapted from (Ma *et al.*, 2013).

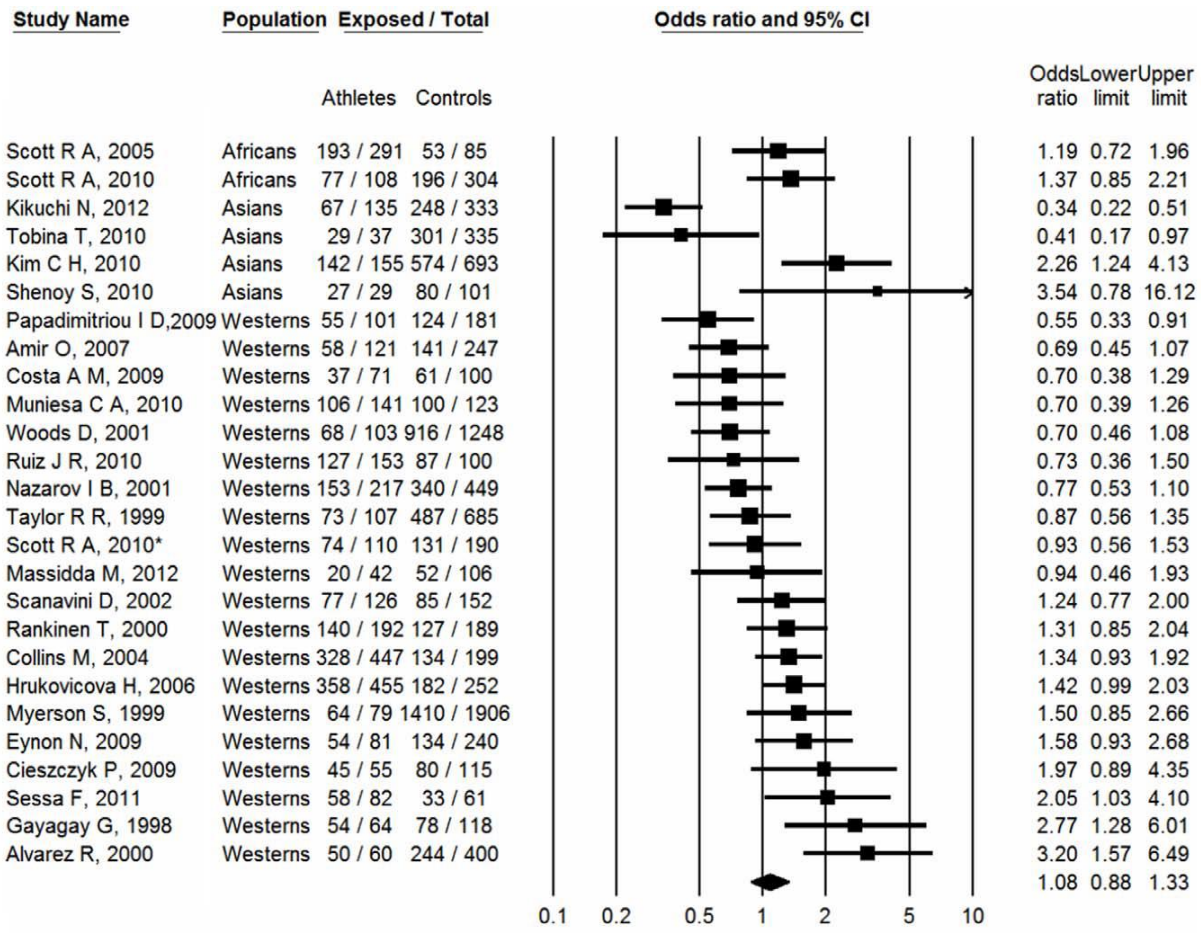
2.3.2 *ACTN3* R577X rs1815739

Alpha-actinins constitute a family of actin-binding proteins necessary to anchor actin filaments to the sarcomeric Z-line (Blanchard *et al.*, 1989; MacArthur & North, 2004). The

α-actinin-3 protein is one such structural component that binds the actin thin filament to the Z line by its distinct N terminal actin binding domain and is expressed almost exclusively in fast glycolytic type II muscle fibres (Beggs *et al.*, 1992; Mills *et al.*, 2001). A functional single nucleotide polymorphism (SNP) in the *ACTN3* gene, located in exon 16 of chromosome 11, has been identified (North *et al.*, 1999) and is characterized by a C > T transition. This results in an arginine codon (R) being replaced by a premature termination codon (X) at the 577 amino acid position. Therefore, RR homozygotes have the fully functioning gene variant, whereas individuals homozygous for the X allele are unable to produce the α-actinin-3 protein and occurs in ~18% of Europeans (Yang *et al.*, 2003). Recently, a dose-dependent effect on α-actinin-3 protein and mRNA expression levels were identified in rodents, where the equivalent to the human RR genotype showed the greatest expression levels (Hogarth *et al.*, 2016).

Furthermore, the R allele has been frequently reported to be associated with elite power and sprint athlete status (Yang *et al.*, 2003; Niemi & Majamaa, 2005; Druzhevskaya *et al.*, 2008; Papadimitriou *et al.*, 2008; Roth *et al.*, 2008; Eynon *et al.*, 2010; Papadimitriou *et al.*, 2016), while the X allele has been associated with endurance performance (Zhang *et al.*, 2003; Niemi & Majamaa, 2005). However, inconsistent observations exist (Saunders *et al.*, 2007). Interestingly, both Yang *et al.* (2003) and Niemi and Majamaa (2005) found a complete absence of the XX genotype in the very best sprinters and low frequencies in good sprinters (~6%), which suggests the importance of the α-actinin-3 protein for high velocity muscular contractions. Additionally, meta-analysis have reported the *ACTN3* R allele association with speed and power performance (OR = 1.21, Figure 10; Ma *et al.*, 2013) and the RR genotype with strength and power (Alfred *et al.*, 2011). Consequently, due to the differences in physical characteristics between rugby union athletes and the general population and the diverse physiological demands within rugby union (Table 1), this genetic marker could

predispose rugby athletes to success or selection at the elite level either through strength and power or endurance characteristics.



Subgroup analysis		Summarized OR (95% CI)	No. of included studies	Heterogeneity test	
				I^2 (%)	*p*
All		1.08 (0.88-1.33)	25	73.91	<0.01
Gender	Male	0.95 (0.67-1.35)	10	77.16	<0.01
	Female	0.59 (0.36-0.98)	2	0.00	0.89
Ethnicity	Westerns	1.10 (0.91-1.33)	20	64.10	<0.01
	Asians	0.94 (0.29-3.03)	4	90.6	<0.01
	Africans	1.28 (0.91-1.81)	2	0.00	0.69
Sport Discipline	Endurance	1.13 (0.89-1.44)	17	64.15	<0.01
	Power	0.91 (0.65-1.28)	13	79.51	<0.01

Figure 10 Meta-analysis of the association between sport performance and *ACTN3* R577X Polymorphism (RR vs. RX+XX). Abbreviation: CI, confidence interval; OR, odds ratio. Adapted from (Ma *et al.*, 2013).

Mechanistically, an α-actinin-3 deficiency might impair the performance of the type II fibres (MacArthur & North, 2007), that are bigger (Broos *et al.*, 2016), are able to contract faster and produce more power than type I muscle fibres (Bottinelli *et al.*, 1996; Gilliver *et al.*, 2009). However, of benefit to certain physiological phenotypes, a higher propensity for aerobic enzymatic activity (porin, COX IV, hexokinase, citrate synthase, succinate dehydrogenase and β-hydroxyacyl CoA dehydrogenase; Seto *et al.*, 2011; Seto *et al.*, 2013) and greater force recovery after fatigue that has been identified in α-actinin-3 deficient mice (Seto *et al.*, 2011). This could indicate that humans with the XX genotype (α-actinin-3 deficiency) might have a greater capacity for recovery from fatiguing exercise. This notion is further corroborated by findings from knockout (KO) mouse models that α-actinin-3 deficiency results in lower muscle mass and strength, longer time to exhaustion and enhanced recovery from fatigue (MacArthur *et al.*, 2008; Chan *et al.*, 2011). Confirmation in human models are beginning to emerge, with one recent investigation showing XX individuals could attain the ventilatory threshold at higher speeds, suggesting that they can sustain higher running speeds at lower exercise intensity (Pasqua *et al.*, 2016). With others showing greater sprint, power and strength abilities for RR individuals (Table 5; Garton & North, 2016).

2.3.3 *FTO* rs9939609

The fat mass and obesity associated (*FTO*) gene is the most investigated gene in the genetics of obesity and has complex molecular mechanisms which are yet to be fully elucidated. Recent genome-wide association studies (GWAS) have identified several common SNP in the human *FTO* gene in association with obesity, body mass index (BMI; Jacobsson *et al.*, 2012), cardiovascular disease and hypertension (Liu *et al.*, 2013; He *et al.*, 2014). These *FTO* SNPs, which are in strong linkage disequilibrium ($r^2 > 0.80$), are located in a cluster on the first intron of chromosome 16 and consequently exhibit similar obesity-related traits (Loos & Yeo, 2014). Thus, within different *FTO* variants, those alleles that have been

positively associated with obesity-related phenotypes are referred to as risk alleles, and those demonstrating a protective effect are referred to as protective alleles. Homozygotes for the minor risk allele consistently demonstrate greater BMI and body mass (3-10 kg) in comparison to protective allele carriers (Table 5; Frayling *et al.*, 2007; Rauhio *et al.*, 2013; Woehning *et al.*, 2013). This greater body mass associated with risk allele carriers is likely to be adipose tissue (Andreasen *et al.*, 2008; Rampersaud *et al.*, 2008; Tanofsky-Kraff *et al.*, 2009; Wing *et al.*, 2009; Liu *et al.*, 2010; Lear *et al.*, 2011; Sonestedt *et al.*, 2011; Luis *et al.*, 2012; Matsuo *et al.*, 2012; Matsuo *et al.*, 2014), although there exist some suggestions of greater fat free mass (FFM) in addition to fat mass (Jess *et al.*, 2008; Sonestedt *et al.*, 2011).

Environmental lifestyle factors (diet and physical activity) have also been investigated for *FTO* gene-environment interactions. Risk allele carriers are more likely to choose a high fat diet when compared to protective allele carriers (Sonestedt *et al.*, 2009; Tanofsky-Kraff *et al.*, 2009; Corella *et al.*, 2011; Lear *et al.*, 2011; Lappalainen *et al.*, 2012; McCaffery *et al.*, 2012; Moleres *et al.*, 2012; Phillips *et al.*, 2012). However, with administration of a high protein diet (25% energy intake) risk allele carriers demonstrated greater reduction in body mass, fat mass and percentage body fat (Zhang *et al.*, 2012), due to a greater reduction in food cravings and appetite suppression than protective allele carriers (Huang *et al.*, 2014). Furthermore, physically active risk allele carriers demonstrate a 30% reduction in the odds of becoming obese and have 36% less body fat compared to inactive individuals (Kilpeläinen *et al.*, 2011). Similarly, data from the HERITAGE Family Study showed that following 20 weeks of endurance training, protective homozygotes exhibited reductions in fat mass, three times greater than risk allele carriers (Rankinen *et al.*, 2010). Interestingly, when comparing normal weight and obese individuals who participate in sport, no differences in *FTO*

variation were observed (P = 0.97), which was contrasted by those not participating (P = 0.02; Muc *et al.*, 2015).

Eynon et al. (2013b) investigated *FTO* rs9939609 in three European cohorts of power (n = 258; 58% elite) and endurance athletes (n = 266; 57% elite) from a variety of sporting disciplines - but identified no associations. This lack of association was likely due to the considerable differences in physiological demand between the varieties of athletic disciplines included, plus further variability in the standard of athlete. Therefore, as rugby includes athletes of remarkably distinct anthropometric and body composition phenotypes, elite rugby provides a unique opportunity to investigate *FTO* in individuals at the extreme upper end of physical fitness (Chapter *2.2*).

2.3.4 *APOE* ε4 rs429358 and rs7412

The *APOE* gene is located on chromosome 19, encodes apolipoprotein E-based peptide (ApoE) and is a candidate marker for risk and severity of mild traumatic brain injury (mTBI). ApoE is a protein that plays a pivotal role in cholesterol metabolism (Weisgraber, 1994) and has been linked to neurobiological function (Teasdale *et al.*, 1997; Laskowitz *et al.*, 2010), specifically, susceptibility to late-onset and sporadic Alzheimer's disease via *APOE* gene polymorphisms (Busjahn *et al.*, 2009). In humans, ApoE is a 299 amino acid protein and has three common isoforms (Apo ε2, ε3, and ε4) which differ by two separate single amino acid changes (both cysteine/arginine). In all three isoforms, the C-terminal domain is largely responsible for lipid binding, whereas the N-terminus is comprised of a four α-helix motif that includes the receptor binding region (Laskowitz & Vitek, 2007). The *APOE* gene has a ε2/ε3/ε4 haplotype derived from two nonsynonymous SNPs (rs429358 and rs7412) within exon 4 and results in three distinct alleles (ε2, ε3, and ε4) with six possible genotypes (*APOE* ε2/ε2, ε2/ε3, ε2/ε4, ε3/ε3, ε3/ε4, and ε4/ε4; Hixson & Vernier, 1990; Bennett *et al.*, 2016).

Carriers of the *APOE* ε4 allele have presented with reduced motor rehabilitation outcomes, poorer neurocognitive outcomes, increased cognitive impairments, amnesia and memory defects following TBI (Lichtman *et al.*, 2000; Crawford *et al.*, 2002; Müller *et al.*, 2009; Noé *et al.*, 2010). Furthermore, multiple meta-analysis have shown an increased risk of poor outcome greater than 6 months post TBI (Zhou *et al.*, 2008; Zeng *et al.*, 2014; Li *et al.*, 2015) with one suggesting that *APOE* ε4 was responsible for up to 64% of the hazardous effect of TBI (Lawrence *et al.*, 2015). Of particular concern for athletic populations, where mTBI is generating increasing interest (McCrory *et al.*, 2013; Raftery, 2013; Fuller *et al.*, 2016; Kemp *et al.*, 2016; Raftery *et al.*, 2016), Lawrence et al. (2015) reported that *APOE* ε4 was accountable for 38% of the 'hazardous influence' towards delaying recovery from mTBI (this analysis included, but was not limited to, concussion data: Figure 11). As concussion is a form of mTBI and is classified as such (Roozenbeek *et al.*, 2013), it will be referred to as mTBI from here on.

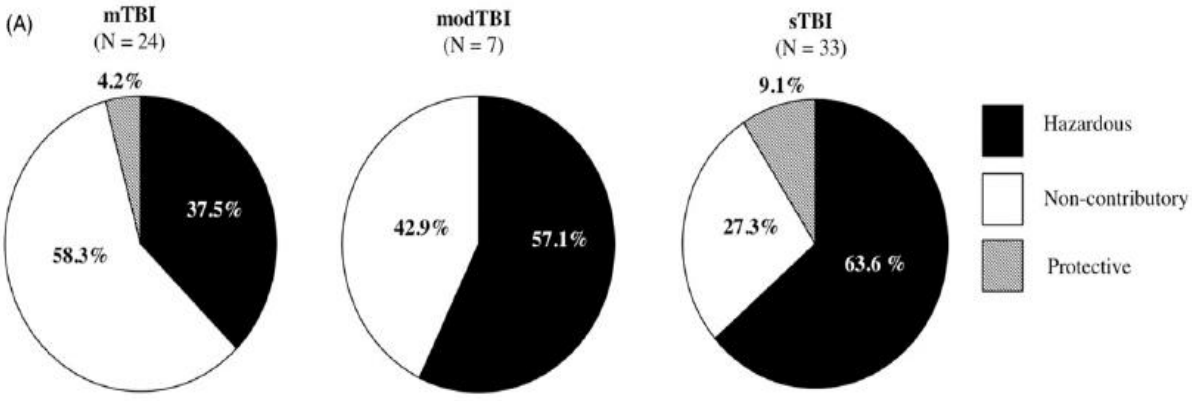


Figure 11 Breakdown of the influence of *APOE* ε4 on TBI outcome at a study-based level on traumatic brain injury (TBI) severity. Adapted from (Lawrence *et al.*, 2015).

There appears to be no association between *APOE* ε4 and self-reported history of sport-related mTBI (Terrell *et al.*, 2008; Tierney *et al.*, 2010) or between prospective mTBI assessment (Kristman *et al.*, 2008). This is not surprising, as large clinical studies show little

APOE ε4 association with 'immediate' severity or morbidity, but instead shows a poorer trajectory towards recovery (Noé *et al.*, 2010; Pruthi *et al.*, 2010; Ponsford *et al.*, 2011). Similar to that of TBI, *APOE* ε4 athletes experience prolonged symptomatic responses to sport-related mTBI (Kutner *et al.*, 2000; Merritt & Arnett, 2016), which have recently been categorised as physical, cognitive, affective and sleep (Merritt & Arnett, 2016). In a sample of 42 college athletes who suffered an mTBI event, Merritt and Arnett (2016) divided participants into two groups, those possessing the ε4 allele (ε4+) and those not (ε4-). Consistently for all post-mTBI symptoms, ε4+ suffered more severe symptoms with a medium effect size for cognitive (Cohen's d = 0.60) and a large effect size for physical symptoms (d = 0.87), more than 3 months post-mTBI (Figure 12). Despite the limited information on *APOE* ε4 and sport related mTBI data, inferences can be made considering

other classifications of TBI to conclude that *APOE* ε4 is a substantial factor in mTBI recovery and as such warrants assessment in elite rugby athletes (Table 5).

At this point it is important to consider the long-term effects of repeated mTBI, first eluded to by Omalu et al. (2005; 2006; 2010) in a series of publications focused on retired American football suicide victims. Here, Omalu et al. identified a pathological condition termed chronic traumatic encephalopathy (CTE). CTE can only be identified by autopsy and is consistent with retired boxers (Geddes *et al.*, 1999). A recent systematic review of CTE cases across various sports identified 153 CTE cases (86% boxers and American football players) and concluded that there is ample evidence to associate mTBI and CTE (Maroon *et al.*, 2015). Furthermore, in 80 cases *APOE* genotype was assessed and the ε4 genotype was

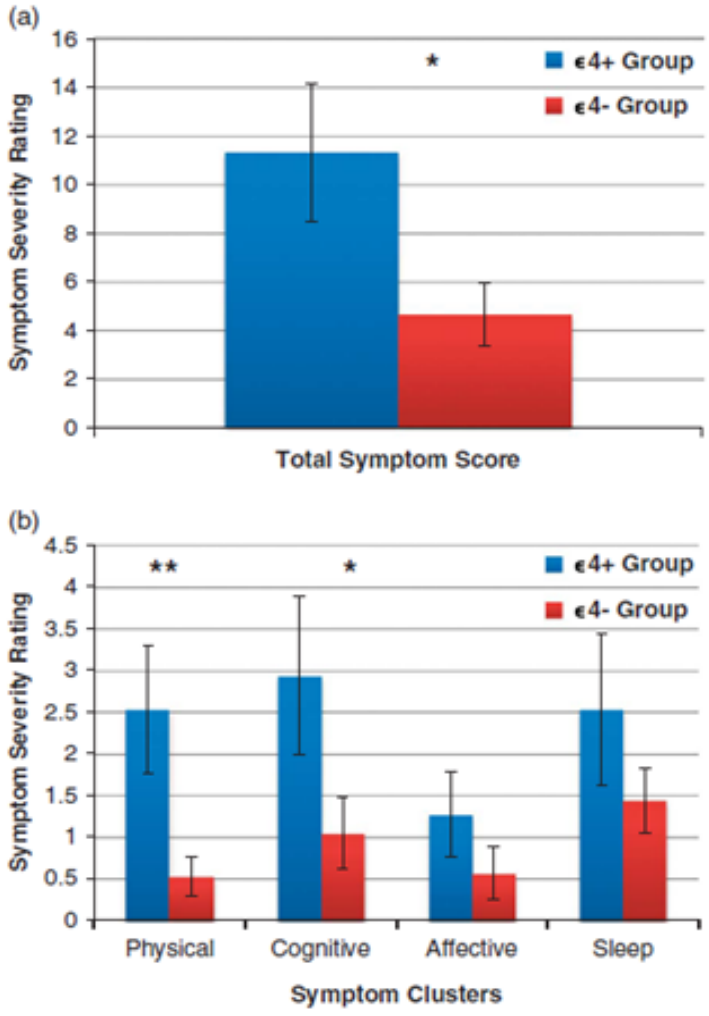


Figure 12 Means and standard errors of each symptom variable are presented in the figure, according to ε4 allele group. Total symptom score comparisons are illustrated in figure **a** and symptom cluster comparisons are illustrated in figure **b**. Adapted from (Merritt & Arnett, 2016).

present in 9.4% of CTE, which is higher than the general population (2.9%), and with relatively normal distribution across other *APOE* genotypes. However, no large studies of CTE and *APOE* ε4 exists (Maroon *et al.*, 2015). It is also likely that older athletes with CTE are often misdiagnosed as having age related neurological decline, therefore no detailed neurological autopsy will be performed. This is possible because 51% of CTE cases are identified in individuals over the age of 60 yr (Maroon *et al.*, 2015). CTE has gathered considerable attention in the rugby scientific community (Raftery, 2013; Calderwood *et al.*, 2015; Stewart *et al.*, 2015; Raftery *et al.*, 2016) and as the identified incidence of mTBI is growing (Chapter *2.1.4*), understanding any genetic predisposition to mTBI could help improve player welfare and management so that players avoid having a post-mortem CTE diagnosis.

Functionally, the *APOE* ε4 allele has also been associated with neurodegenerative cascade subsequent to TBI, the severity of axonal injury in mouse models (Sabo *et al.*, 2000; Hartman *et al.*, 2002; Bennett *et al.*, 2013) and more likely to show deposition of β-amyloid in brain tissue following head injury (Nicoll *et al.*, 1995; Teasdale *et al.*, 1997). Furthermore, in the intensive care setting following TBI, ApoE ε4 isoform have increased systemic CNS inflammatory responses (Lynch *et al.*, 2003) and the *APOE* ε4 polymorphism has been associated with increased systemic inflammatory responses (Moretti *et al.*, 2005). The cellular mechanisms by which the molecular processes of *APOE* isoforms' differ are not fully understood. However, owing to the role of ApoE peptides, inferences can be made in relation to neurophysiological inflammatory response.

In microglial cell cultures exposed to ApoE, phosphorylation of c-Jun N-terminal kinases (JNK) leads to suppression of important inflammatory factors such as TNF-α, IL -6 and IL-12 (Hidding *et al.*, 2002; Pocivavsek *et al.*, 2009a). ApoE induces a reduction of JNK

phosphorylation and suppresses nitric oxide synthase synthesis (Pocivavsek *et al.*, 2009a) - a key neurotransmitter. Furthermore, this cellular interaction appears to be mediated by lipoprotein receptor-related protein (LRP), of which show ApoE isoforms express specific binding (ε3 binds with greater affinity than ε4; Pocivavsek *et al.*, 2009b; Bell *et al.*, 2012b). These data suggest that ApoE isoforms may mediate microglial immune response which could be triggered by cellular trauma, such as TBI. Moreover, animal models have shown that ApoE ε4 isoforms are associated with Mitogen-activated protein kinase (MAPK) signalling pathways (regulator of, proliferation, expression, differentiation and apoptosis; Maezawa *et al.*, 2006). This is interesting given the recent finding that *APOE* ε4 increased trauma-induced-early-apoptosis via a reduction in potassium current in a neuronal/glial cell cultures, resulting in an increase of intracellular calcium (Chen *et al.*, 2015). For a more comprehensive review of the proposed biological mechanisms see Gokhale & Laskowitz (2013). The current understanding of *APOE* ε4 is continuing to grow and these *in vitro* and mouse models further support the notion of impaired recovery following mTBI. Therefore, investigating if these 'risk' individuals have been filtered out before competing at the elite level of competitive rugby is a necessary step towards understanding the molecular bases of mTBI in rugby union.

2.3.5 *COL5A1* rs12722 and rs3196378

Probably the most explored gene regarding tendon and ligament injuries is the *COL5A1* gene, which encodes for a minor fibrillar collagen protein (Hildebrand *et al.*, 2004). Collagen is the primary structurally connective tissue protein of the extra cellular matrix (Figure 13) that regulates fibrogenesis through its fibril structure and diameter (Birk *et al.*, 1990; Chanut-Delalande *et al.*, 2004; Wenstrup *et al.*, 2006). Two amino acid components (collagen type V and type I fibrils) co-polymerise to form heterotypic fibres. The major collagen type V isoform comprises two α-1-(V) chains, encoded by the *COL5A1* gene, one α-2-(V) chain

encoded by the *COL5A2* gene (Wenstrup *et al.*, 2004; Malfait *et al.*, 2010) which forms between 1-3% of total collagen content (Birk *et al.*, 1990; Chanut-Delalande *et al.*, 2004; Wenstrup *et al.*, 2004; Sun *et al.*, 2011). Mutations in the *COL5A1* gene have been identified in Ehlers-Danlos syndrome, a disease characterised by

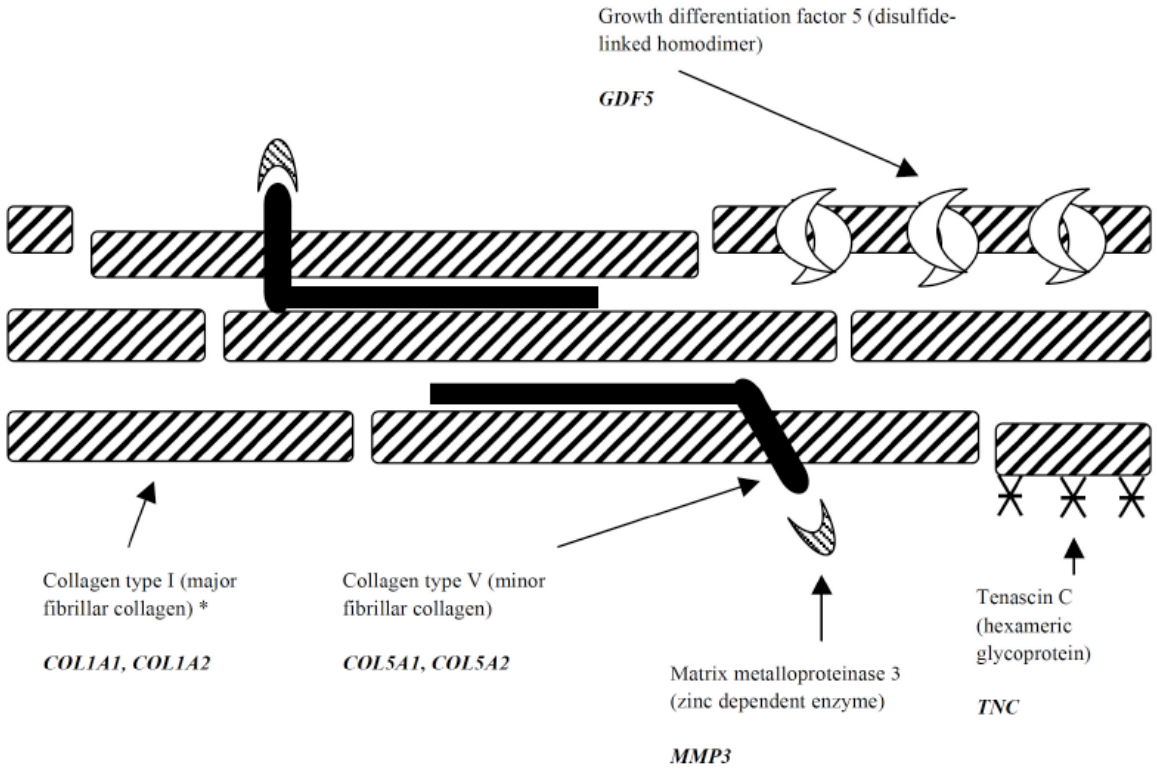


Figure 13 Major microstructural components of tendons associated with tendon pathologies/musculotendinous range of motion, identifying related genes. Adapted from Foster et al. (2012).

joint hypermobility, laxity and muscle hypotonia (Beighton *et al.*, 1998), disrupt collagen type V organisation and can affect the assembly of other collagens in the extra cellular matrix (Zoppi *et al.*, 2004). This results in irregularly large collagen fibrils located within connective tissue (Vogel *et al.*, 1979) and is attributed to a reduced synthesis of collagen type V (Malfait & De Paepe, 2005; Sun *et al.*, 2011).

Two common *COL5A1* gene SNPs (rs12722 and rs3196378), located in the 3′ untranslated region (3′ UTR), are the subject of the present chapter because of their association with tendon (September *et al.*, 2009) and ligament (rs12722; Posthumus *et al.*, 2009a) pathology and their sequence proximity (Figure 14). Both rs12722 and rs3196378 were associated with tendinopathy in Australian Caucasians, but only the former was associated with South African Caucasians (September *et al.*, 2009). This is interesting considering both SNP were in linkage disequilibrium (D′ ≥ 0.67; September *et al.*, 2009; Laguette *et al.*, 2011).

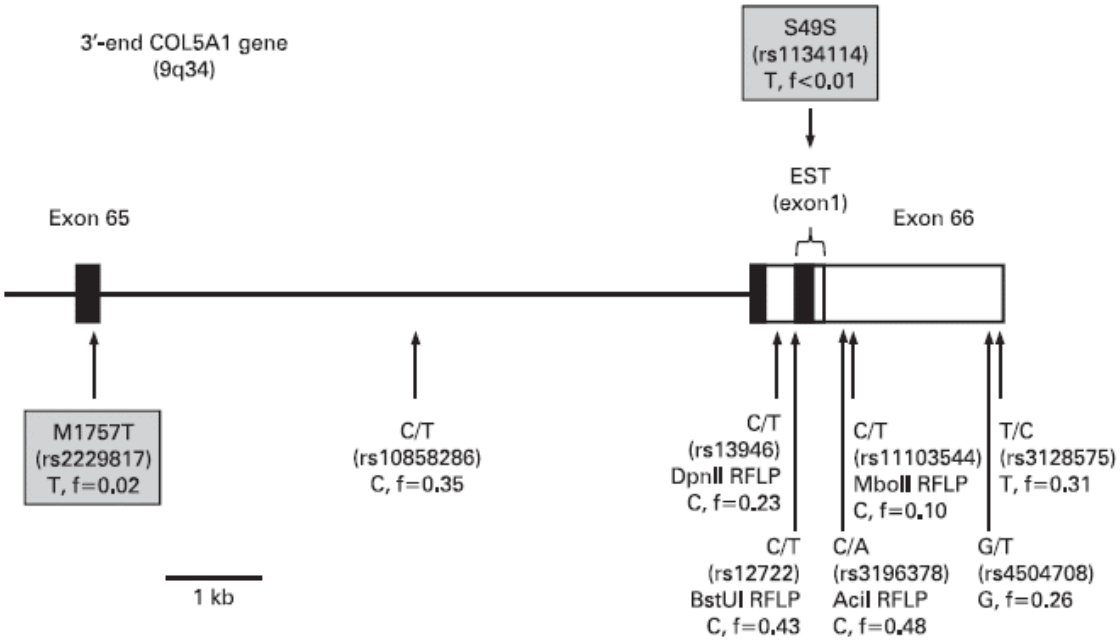


Figure 14 A schematic representation of the terminal exons (rectangles) and intron (horizontal lines) boundaries of the 39-end of the *COL5A1* gene. The translated regions of the exons are solid whereas the untranslated region (UTR) of exon 66 is clear. All the information used to construct this figure was obtained from databases hosted by the NCBI (*http://www.ncbi.nlm.nih.gov/*) and Ensembl (*http://www.ensembl.org/index.html*). Adapted from (September *et al.*, 2009)

Furthermore, rs12722 has also been previously associated with flexibility (Collins *et al.*, 2009; Lim *et al.*, 2015), anterior cruciate ligament injury (Posthumus *et al.*, 2009b; Altinisik *et al.*, 2015; O'Connell *et al.*, 2015) and Achilles tendinopathy (Mokone *et al.*, 2006), but not patellar tendon properties (Foster *et al.*, 2014). In these studies, the minor CC genotype was shown to be overrepresented in the respective asymptomatic controls, suggesting a

protective role of the C allele against injury. Considering the high frequencies of tendon and ligament inquiries in elite rugby (Williams *et al.*, 2013; King *et al.*, 2014; Fuller *et al.*, 2015a; Fuller *et al.*, 2016), assessing these specific genetic variants may be of use to help improve management of individual player injury risk (Table 5).

Some possible mechanisms have been proposed to explain the association of *COL5A1* gene variants and soft tissue injury (Laguette *et al.*, 2011; Abrahams *et al.*, 2013). Laguett et al. (2011) have shown that the *COL5A1* 3′ UTR - wHere both rs12722 and rs3196378 are situated - affects miRNA stability. For both SNPs, the alleles associated with greater soft-tissue injury risk were associated with greater Hsa-miR-608 stability, which in turn may alter the Col5α1 protein secondary structure - proposed to play a role in type V collagen production (Abrahams *et al.*, 2013). This would suggest that C/T allele differences at rs12722 may alter the co-polymerisation of collagen type V and type I fibrils. However, to date, this has not been demonstrated experimentally and exactly how this may translate into functional properties is currently unknown. Assessment of these *COL5A1* genetic markers, in combination with others yet to be identified, might provide a useful tool in rugby for individualising training load and mode to reduce incidence of injury. Given that the *COL5A1* variants are associated with injury phenotypes, it is plausible that at the elite level of rugby union, risk allele individuals may be selected out on the bases of repeated injuries which may be reflected in the quantification of genetic status.

Table 5 Thesis candidate gene variants and associated phenotypes.

Gene variant	Phenotype	Reference
***ACE* I/D**	I allele associated with greater endurance capacity and physical performance.	(Puthucheary *et al.*, 2011; Jang & Kim, 2012; Ma *et al.*, 2013)
ACTN3	R allele associated with greater power performance and associated phenotypes	(Ma *et al.*, 2013; Orysiak *et al.*, 2015; Papadimitriou *et al.*, 2016)
FTO	A allele associated with greater (fat) mass, T allele associated with obesity protection	(Frayling *et al.*, 2007; He *et al.*, 2014)
***APOE* ε4**	Carriage of the ε4+ allele associated with poorer recovery following mTBI	(Kutner *et al.*, 2000; Ponsford *et al.*, 2011; Merritt & Arnett, 2016)
COL5A1	C allele associated with soft tissue injury protection and resistance to muscle cramping	(Posthumus *et al.*, 2009b; September *et al.*, 2009; O'Connell *et al.*, 2013; Collins *et al.*, 2015)

2.3.6 Thesis Aims and Hypothesis

The overarching aim of the present thesis was to investigate whether elite rugby athletes, as part of the RugbyGene project, differed in terms of genetic variation from a control group and whether athletes in specialized playing positions similarly differed. As such, the specific aims of the present research project were;

- To recruit a large biobank of elite rugby union athletes for the purpose of evaluating the molecular genetic components of elite rugby athlete status and to investigate the molecular underpinnings of the physiological and anthropometric differences that exists between elite rugby playing position.

- To investigate *ACE* I/D and *ACTN3* R577X genotype distribution in elite rugby athletes. Whereby it was hypothesized that the *ACTN3* R allele and the *ACE* I allele would be more frequent in rugby athletes than controls. It was further hypothesized that *ACTN3* XX and *ACE* II genotypes would be underrepresented in backs compared to forwards, due to differences in overall work-to-rest ratio and differing requirements for high maximum sprinting speed.

- To investigate if associations of *FTO* rs9939609 genotype differ between elite rugby athletes and a control population, and/or between playing positions. Based on prior data in obese populations, it was firstly hypothesised that the rs9939609 risk (A) allele would be overrepresented in playing positions typically requiring greater body and muscle mass, while the protective (T) allele would be more common in positions requiring a lean phenotype.

- To quantify the 'at risk' *APOE* ε4 carriers in elite rugby athletes and to investigate if *APOE* genotypes differed between elite rugby athletes and a control population. Based on the published *APOE* ε4/ε4 association with poorer outcome following brain injury, it was hypothesised that the ε4/ε4 genotype and ε4+ would be underrepresented in elite rugby athletes compared to controls.

- To investigate if associations of *COL5A1* rs12722 and rs3196378 genotype and allele frequencies differed between elite rugby athletes and a control population. It was hypothesised that the *COL5A1* rs12722 and rs3196378 protective C allele and CC genotype would be overrepresented in elite rugby athletes compared to controls.

- Finally, a total genotype score (TGS) algorithm will be applied to assess the polygenic effect, of the gene variants examined in the first four experimental chapters of the present thesis, for all RU athletes, forwards and backs, compared to controls.

Chapter 3

Experimental methods

3.1 Methods

3.1.1 Participants

For sports genomics research to be truly relevant to the preparation and management of elite competitors, the athletes from which the initial data are derived must themselves be considered elite. In the context of rugby union, the definition of 'elite' is proposes as athletes competing in the highest competitive league of a 'Tier 1' rugby nation (Regulation 16, *www.worldrugby.org*) - (International Rugby Board, 2004). Given the evolving nature of elite rugby, the era in which athletes competed at an elite level also needs to be defined. Rugby union has changed dramatically in the > 100 years of its existence and that change has certainly continued significantly since the sport turned professional ~20 years ago. Nevertheless, it is proposed that 1995 onwards is a playing era inclusion criterion that can sensibly be justified. Geographic ancestry is another important consideration for case-control and genotype-phenotype association study designs and therefore analysis of molecular genetic markers should preferably be performed on athletes from a well-defined geographic ancestry cluster – in the present thesis, Caucasians of European descent. No population stratification was evident between UK and SA population within the present sample ($P > 0.05$), genotypic frequencies are presented in appendix 6) A more difficult aim to achieve would be to recruit large numbers of players from all geographic ancestry clusters commonly found in rugby union, although this would be a very powerful approach scientifically and a long term goal of the RugbyGene Project.

Ethical approval was granted by Manchester Metropolitan University (MMU), University of Glasgow, University of Cape Town and Northampton University ethics committees and complies with the Declaration of Helsinki (2013), as part of the RugbyGene project. Sample size and participant details will differ for each experimental chapter due to increasing participant recruitment throughout the project time-frame, genetic data availability and as

participants become eligible for inclusion in the elite cohort. Therefore, precise sample numbers and anthropometric data will be specified in each experimental chapter (*4*, *5*, *6*, *7* and *8*).

3.1.2 Sample collection

The majority of the blood (~70% of all samples), saliva (~25%) and buccal swab samples (~5%) that were obtained by the author (as lead researcher on the RugbyGene project research team), via the following protocols. A 5 mL blood sample was drawn from a superficial forearm vein, by a trained phlebotomist, into an ethylenediamine tetra acetic acid (EDTA) anticoagulant treated tube (BD Vacutainer Systems, Plymouth, UK) and stored in 2 mL sterile tubes (Eppendorf AG, Hamburg, Germany) at -20˚C until processing. Given that ~24% of people have a phobia of needles (Taddio *et al.*, 2012), alternative non-invasive sample collection methods were also used. Saliva samples were collected into Oragene DNA OG-500 collection tubes (DNA Genotek Inc., Ontario, Canada) according to the 5 step manufacturer's protocol (Figure 15). Greater than 30 min abstinence from food and drink, participants drooled into the collection tube until the amount of liquid saliva (not bubbles/foam) reaches a fill line printed on the collection tube, the screw cap was tightened to secure the tube and was gently invert and evert for ~5 seconds (Figure 15). These samples were then stored at room temperature until processing, followed by long-term storage at -20˚C. For buccal cell collection, following a minimum 1-hour abstinence from food and drink, sterile buccal swabs (Omni swab, Whatman, Springfield Mill, UK) were rubbed against the buccal mucosa of the cheek for approximately 30 s. A second swab was collected from the opposite cheek. Tips were ejected by firmly pressing the plunger at the end of the handle into sterile 2 mL tubes and stored at -20˚C until processing. All collection tubes, from all methods, were coded and labelled to ensure participant anonymity in accordance with the Human Tissue Act (2004).

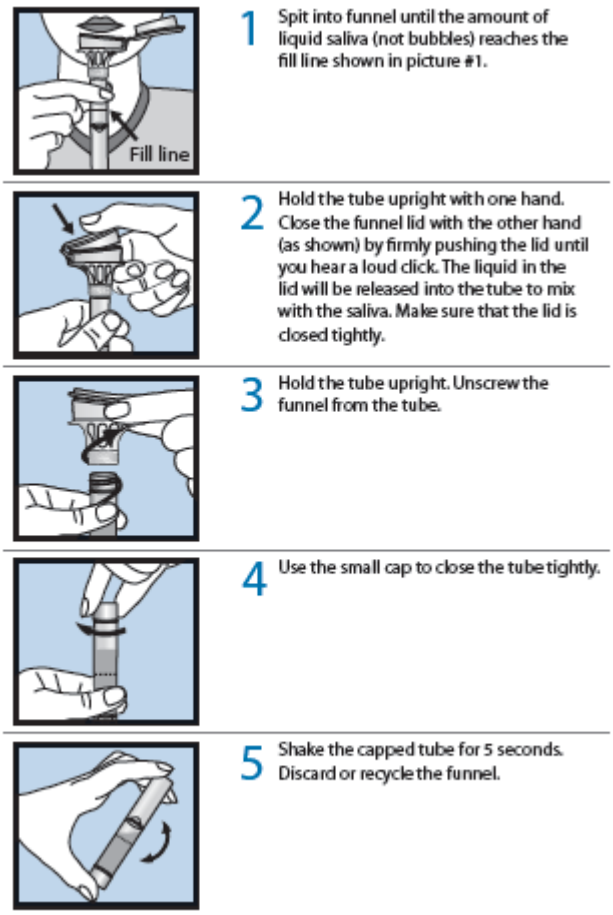


Figure 15 Oragene DNA collection kit procedure (*http://www.dnagenotek.com/ROW/support/ciOG500.html*).

3.1.3 DNA extraction

DNA isolation and genotyping were performed in MMU (by the author), University of Glasgow, University of Cape Town (DNA isolation only) and University of Northampton laboratories. There are some differences between protocols summarized below. There was 100% agreement among reference samples genotyped in the three genotyping centres, i.e. Glasgow, Northampton and MMU laboratories (verified by the author). The majority of samples were processed and genotyped in the MMU Sports Genomics laboratory by the author. Genotype calling was 100% successful for all polymorphisms in the athlete samples and for the majority of the control samples, however 10 of the 566 control samples for

rs3196378 were unsuccessful despite repeated attempts and these were genotyped in the Glasgow lab.

At MMU (performed by the author) and Glasgow, DNA isolation was performed using the QIAamp DNA Blood Mini kit and standard spin column protocol, following the manufacturer's instructions (Qiagen, West Sussex, UK). 200 µL of whole blood/saliva, or one buccal swab, were combined with protease and incubated at 56°C for 10 min. Ethanol (96%) was added and the mixed fluid was centrifuged at 8000 rpm for 60 s leaving the DNA sample bound to the spin column silica-gel membrane (Appendix 3). Wash buffers were passed through the sample for the removal of proteins, nucleases and other impurities. Finally, a low-salt pH-balanced elution buffer was used to provide a 100 µL solution containing isolated DNA, which was stored at 4°C until further analysis.

In Cape Town, DNA was isolated from whole blood using a different protocol (Lahiri & Nurnberger, 1991). Each sample was combined with lysis buffer, nuclei were pelleted by centrifugation and re-suspended in a high salt buffer. DNA was further precipitated following protein digestion, 100% ethanol was added and the sample was centrifuged, washed with 70% ethanol and dried. DNA hydration buffer was added and samples were stored at -20ºC until subsequent analysis. Genotyping of DNA isolated in Cape Town was performed in Glasgow.

At Northampton, DNA was isolated from whole blood using Flexigene kits (Qiagen). Each sample was combined with lysis buffer, nuclei were pelleted by centrifugation and re-suspended in protease-containing denaturation buffer. DNA was further precipitated following protein digestion, isopropanol was added and the sample was centrifuged, washed with 70% ethanol and then dried. DNA hydration buffer was added and samples were stored at -20ºC until subsequent analysis.

3.1.4 Genotyping

Genotyping of selected SNP assays (Chapter 2.3) were performed slightly differently depending on genotyping centre and are described as follows. In the Glasgow laboratory, genotyping was performed by 10 µL Genotyping Master Mix (Applied Biosystems, Paisley, UK), 1 µL SNP-specific TaqMan assay (Applied Biosystems), 6 µL nuclease-free H_2O and 3 µL DNA solution (~9 ng DNA) were added per well. In the Northampton laboratory, the author performed all genotyping by combining 10 µL of Genotyping Master Mix, 8 µL H_2O, 1 µL assay mix with 1 µL of purified DNA (~10 ng). In both laboratories, PCR was performed using a StepOnePlus™ Real-Time detector (Applied Biosystems). Denaturation began at 95°C for 10 min, with 40 cycles of incubation at 95°C for 15 s then annealing and extension at 60°C for 1 min. Initial analysis was performed using StepOnePlus™ software version 2.3 (Applied Biosystems, *https://www.thermofisher.com/uk/en/home/technical-resources/software-downloads/StepOne-and-StepOnePlus-Real-Time-PCR-System.html*; Figure 16).

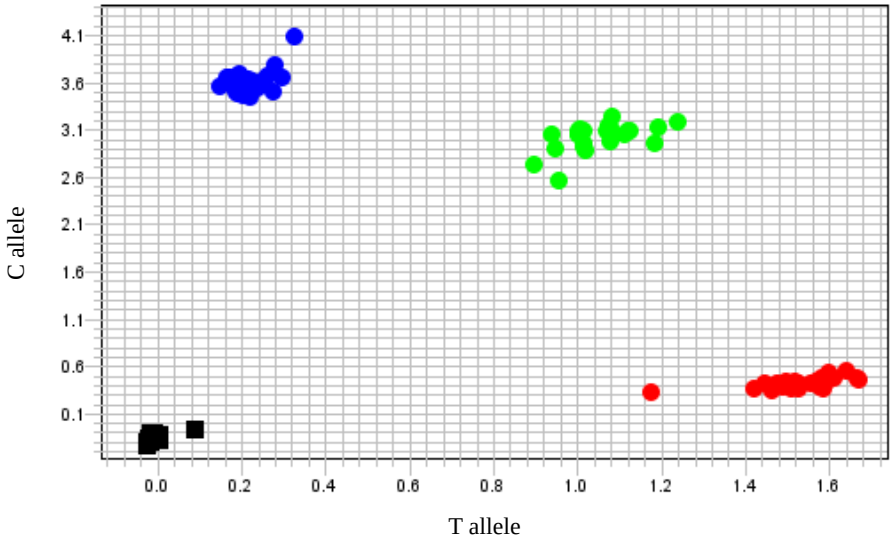


Figure 16 Example allelic discrimination plot for *COL5A1* rs12722 obtained using the StepOnePlus™ Real-Time PCR System.

At MMU, the author (aided by other researchers of the MMU sports genomics team) genotyped all samples by combining 5 µL Genotyping Master Mix, 4.3 µL H_2O, 0.5 µL assay mix (unless otherwise described in experimental chapters) and 0.2 µL of purified DNA (~9 ng), for samples derived from blood and saliva. For DNA derived from buccal swabs, 5 µL Genotyping Master Mix was combined with 3.5 µL H_2O, 0.5 µL assay mix and 1 µL DNA solution (~9 ng DNA). Either a Chromo4 real-time system (Bio-Rad, Hertfordshire, UK; Figure 17) or a StepOnePlus™ Real-Time detector was used. Denaturation began at 95°C for 10 min, with 40 cycles of incubation at 95°C for 15 s then annealing and extension at 60°C for 1 min. Initial genotyping analysis was performed using Opticon Monitor software version 3.1 (*http://www.bio-rad.com/en-uk/sku/soft-om-sw-opticon-monitor-software*) or StepOnePlus™ software version 2.3 (*https://www.thermofisher.com/uk/en/home/technical-resources/software-downloads/StepOne-and-StepOnePlus-Real-Time-PCR-System.html*). There was 100% agreement within duplicates of all samples, in all genotyping centres.

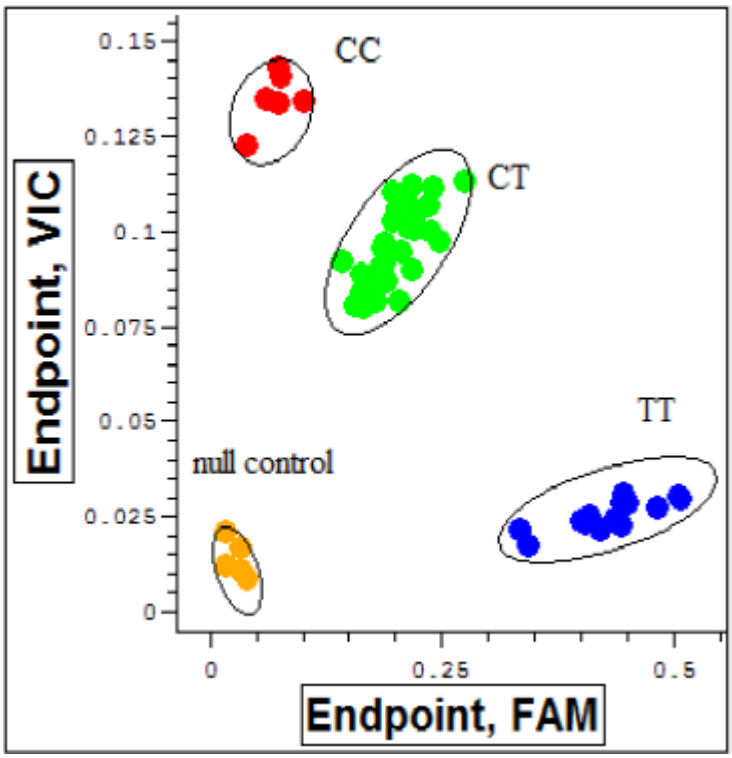


Figure 17 Example allelic endpoint plot for *COL5A1* rs12722 obtained using the Bio-Rad™ Chromo4 real-time system. Endpoint VIC representing the rs12722 C allele and Endpoint FAM representing the rs12722 T allele.

Each experimental chapter will have an abbreviated version of the above DNA isolation and genotyping information with additional content in relation to the specific method required for each assay in each experiment. For example, in chapter *4.1.2* the *ACE* gene insertion/deletion assay is used and requires three separate primers and probes (Koch *et al.*, 2005). The genotyping section of each experimental chapter will include this, and other, additional and differing details.

3.1.5 Positional specification

To assess genotype and allele frequencies within rugby union (RU) players, athletes were allocated to subgroups; forwards (props, hookers, locks, flankers, number eights) and backs (scrum halves, fly halves, centres, wings, full backs). Also, due to diverse game demands (Figure 18; Roberts *et al.*, 2008; Cahill *et al.*, 2013; Quarrie *et al.*, 2013; Jones *et al.*, 2015) and physiological quantities within rugby union (Table 1; Appleby *et al.*, 2012; Sedeaud *et al.*, 2012; Fuller *et al.*, 2013; Sedeaud *et al.*, 2013; Smart *et al.*, 2013; Barr *et al.*, 2014), athletes were further divided into positional groups according to their similar movement patterns (Cahill *et al.*, 2013) and further physiological differences (Table 1), front five (props, hookers, locks), back row (flankers, number eights), half backs (scrum halves, fly halves), centres and back three (wings and full backs).

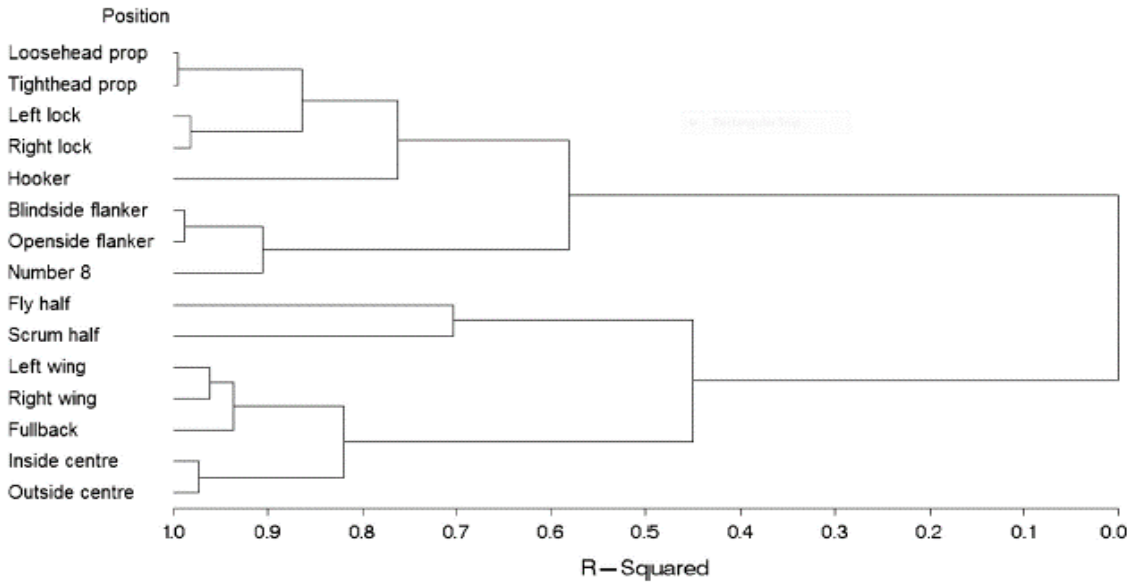


Figure 18 Game demand cluster analysis of rugby union playing positions. Adapted from (Quarrie *et al.*, 2013).

3.1.6 Statistics

From the statistical limitations identified in previous team sport genomic investigations (Chapter *2.2.7*), the present thesis will aim to adhere to the suggested statistical recommendations. Specifically the use of the HWP (Chapter *2.2.7.1*), multiple testing correction procedures (Chapter *2.2.7.2*) and estimating the effect size of any associations (Chapter *2.2.7.3*). As such, SPSS for Windows version 22 (SPSS Inc., Chicago, IL) software was used to conduct Pearson's Chi-square (χ^2) tests to compare genotype and allelic frequencies between athletes and controls, and between positional subgroups. Where appropriate, CubeX online software (*http://www.oege.org/software/cubex*) was used to determine haplotype frequencies and linkage disequilibrium statistics (Gaunt *et al.*, 2007). All P values generated from null-hypothesis testing were subjected to Benjamini-Hochberg corrections (BH; Benjamini & Hochberg, 1995) to control false discovery rate (Chapter *2.2.7.2*) and corrected probability values are reported throughout each experimental chapter (Chapter 4-8). Odds Ratios were calculated to estimate effect size using the highly recommended (Lukic, 2003) MedCalc online statistics calculator (*https://www.medcalc.org/calc/odds_ratio.php*). Additionally, receiver operating

characteristic (ROC) curves were used to estimate the sensitivity of a total genotype score (comprising the earlier mentioned SNPs, Chapter *2.3*) to detect differences between backs and forwards (Zweig & Campbell, 1993), similar to that used by Ben-Zaken et al. (2015). Alpha was set at 0.05.

Chapter 4

Association of ACTN3 R577X but not ACE I/D gene variants with elite rugby union player status and playing position

This chapter is published as:

Heffernan, S. M., Kilduff, L. P., Erskine, R. M., Day, S. H., McPhee, J. S., McMahon, G. E., ... & Cook, C. J. (2016). Association of *ACTN3* R577X but not *ACE* I/D gene variants with elite rugby union player status and playing position. *Physiological Genomics*, 48(3), 196-201.

4.1 Association of *ACTN3* R577X but not *ACE* I/D gene variants with elite rugby union player status and playing position

4.1.1 Introduction

Rugby is an intermittent team sport comprised of diverse playing positions, each with different physiological, anthropometric and technical attributes (Deutsch *et al.*, 2007; Roberts *et al.*, 2008; Cahill *et al.*, 2013) including two distinct subgroups: forwards and backs. Recently, global positioning system (GPS) tracking and time-motion analysis have been used to estimate the physical demands of rugby athletes and compare forwards and backs during high-level match play (Roberts *et al.*, 2008; Cahill *et al.*, 2013; Jones *et al.*, 2015). Backs travelled 12% greater total distance (6545 m versus 5850 m), achieved maximum speeds 16% faster (30.4 $km \cdot h^{-1}$ versus 26.3 $km \cdot h^{-1}$) and engaged in over four times (58% versus 13%) high-intensity running activities (> 5.0 $m \cdot s^{-1}$), as a proportion of total activity (Roberts *et al.*, 2008; Cahill *et al.*, 2013) compared to forwards. These data suggest a more sprint-oriented metabolic demand in backs compared to forwards. Furthermore, due to the complexities of forward play, forwards performed sixfold more (9.9%) high-intensity static exertion activities (rucks, mauls, scrums and line-outs) than backs (1.6%) and spent 19.8% more time running above 80% of their maximal speed (Roberts *et al.*, 2008; Cahill *et al.*, 2013, respectively). This implies that forwards, although often of higher body mass, (Fuller *et al.*, 2013) are more likely to benefit from fatigue-resistant physiological qualities than backs. Accordingly, Deutsch *et al* (2007) showed that forwards had a notably higher work-to-rest ratio than backs (1:7 and 1:22, respectively). Given that the roles of backs and forwards differ significantly in terms of physiological demands, these differences may be reflected in distinct genetic characteristics.

The two most studied gene variants in exercise genomics (*ACE* I/D and *ACTN3* R577X polymorphisms) have recently been considered in meta-analyses. Ma *et al* (2013) reported that *ACE* II genotype was associated with physical performance (odds ratio (OR) 1.23), especially endurance performance (OR = 1.35). Furthermore, *ACTN3* RR genotype was associated with speed and power performance (OR 1.21; Ma *et al.*, 2013), supported elsewhere (Alfred *et al.*, 2011). More extensive information regarding *ACE* I/D and *ACTN3* R577X polymorphisms is available (Puthucheary *et al.*, 2011; Eynon *et al.*, 2013a). Due to differences in physical characteristics between rugby athletes and the general population and the diverse physiological demands within rugby, these genetic markers could predispose athletes to success or specific roles at the elite level.

One recent paper examined *ACE* I/D genotype frequency distribution in young, non-elite rugby athletes. *ACE* I/D genotype frequencies did not differ between forwards and backs, with no control group included (Bell *et al.*, 2010). Despite this result, because presence of the *ACE* I allele is associated with lower circulating (Rigat *et al.*, 1990; Almeida *et al.*, 2010) and tissue (Danser *et al.*, 1995) ACE enzyme activity, I allele carrying individuals generate less vasoconstrictive angiotensin II (Dzau, 1988a; Munzenmaier & Greene, 1996) and reduced degradation of vasodilating kinins (Dietze & Henriksen, 2008). As such, the ACE enzyme, the main active product of the rennin-angiotensin system (RAS; Erdös & Skidgel, 1987) is the system responsible for control and regulation of blood pressure/volume and exists in adipose tissue (Jonsson *et al.*, 1994), human myocardium (Dzau, 1988b), and skeletal muscle (Reneland & Lithell, 1994). Therefore, I allele carriers may have more advantages hemodynamic flow to working muscles and as such, might be beneficial for elite rugby performance. Bell et al. (2012c) also investigated *ACTN3* R577X in 102 young male rugby union athletes and reported no association, despite some tendencies for the R allele to be more frequent in backs or subgroups of backs. Studying elite athletes would be better able

to answer the question whether these genetic variants are associated with elite status and playing position in rugby.

Therefore, the purpose of the present study was to investigate whether elite rugby athletes and a control group differed in terms of *ACE* I/D and *ACTN3* R577X genotype distribution, and whether athletes in specialized playing positions similarly differed. It was hypothesized that the *ACTN3* R allele and the *ACE* I allele would be more frequent in rugby athletes than controls. It was further hypothesized that *ACTN3* XX and *ACE* II genotypes would be underrepresented in backs compared to forwards, due to differences in overall work-to-rest ratio and differing requirements for high maximum speed.

4.1.2 Method

Participants

Ethical approval was granted by Manchester Metropolitan University (MMU), University of Glasgow, University of Cape Town and Northampton University ethics committees and complies with the Declaration of Helsinki (2013). As part of the RugbyGene project, elite Caucasian male rugby athletes (n = 427; mean (standard deviation) height 1.85 (0.07) m, mass 101 (14) kg, age 29 (7) years) including 71.2% British, 17.2% South African, 7.1% Irish and 4.5% of other nationalities were recruited, having given written informed consent. Caucasian controls (61% male; n = 710; height 1.73 (0.10) m, mass 74 (13) kg, age 29 (16) years) included 89.6% British, 8.9% South African, 0.7% Irish and 0.8% of other nationalities. Of the athletes, 53.4% had competed at international level for a “High Performance Union” (Regulation 16, worldrugby.org). International status was confirmed as of 1 January 2015.

Sample collection

Blood (n = 796 of all samples), saliva (n = 285) or buccal swab samples (n = 57) were obtained via the protocols detailed in Chapter *3.1.2*. Blood sample was drawn from a superficial forearm vein, into an EDTA tube and stored in 2 mL sterile tubes at -20˚C until processing. Saliva samples were collected into Oragene DNA OG-500 collection tubes according to the manufacturer's protocol and stored at room temperature until processing. Omni swab Sterile buccal were rubbed against the buccal mucosa of the cheek for approximately 30 s. Tips were ejected into sterile tubes and stored at -20˚C until processing.

DNA isolation

DNA isolation and genotyping were performed in the MMU, University of Glasgow, University of Cape Town (DNA isolation only) and University of Northampton laboratories. There are some differences between protocols summarized in Chapter *3.1.3*, however there was 100% agreement among reference samples genotyped in the three genotyping centres, i.e. Glasgow, Northampton and MMU laboratories. The majority of samples were processed and genotyped in the MMU laboratory. Genotype calling was successful for both variants in all samples.

Genotyping

Genotyping in the Glasgow laboratory was performed on *ACTN3* (rs1815739) and an *ACE* tag SNP (rs4341) in perfect linkage disequilibrium with *ACE* I/D in Caucasians (Glenn *et al.*, 2009) and Asian (Tanaka *et al.*, 2003). Briefly, 10 µL Genotyping Master Mix (Applied Biosystems, Paisley, UK), 1 µL SNP-specific TaqMan assay (Applied Biosystems), 6 µL nuclease-free H_2O and 3 µL DNA solution (~9 ng DNA) were added per well. In the Northampton laboratory, genotyping was performed for *ACTN3* R577X (rs1815739) by

combining 10 µL of Genotyping Master Mix, 8 µL H_2O, 1 µL assay mix with 1 µL of purified DNA (~10 ng). In both laboratories, PCR was performed using a StepOnePlus™ real-time detector (Applied Biosystems). Briefly, denaturation began at 95°C for 10 min, with 40 cycles of incubation at 95°C for 15 s then annealing and extension at 60°C for 1 min. Initial analysis was performed using StepOnePlus™ software version 2.3 (Applied Biosystems). There was 100% agreement within duplicates of all samples.

At MMU, samples were genotyped for *ACTN3* R577X (rs1815739) by combining 5 µL Genotyping Master Mix, 4.3 µL H_2O, 0.5 µL assay mix and 0.2 µL of purified DNA (~9 ng), for samples derived from blood and saliva. For DNA derived from buccal swabs, 5 µL Genotyping Master Mix was combined with 3.5 µL H_2O, 0.5 µL assay mix and 1 µL DNA solution (~9 ng DNA). Either a Chromo4 real-time system (Bio-Rad, Hertfordshire, UK) or a StepOnePlus™ was used. Briefly, denaturation began at 95°C for 10 min, with 40 cycles of incubation at 95°C for 15 s then annealing and extension at 60°C for 1 min. Initial genotyping analysis was performed using Opticon Monitor software version 3.1 (Bio-Rad) or StepOnePlus™ software version 2.3. Duplicates of all samples were in 100% agreement. For *ACE* I/D at MMU, 5 µL of Genotyping Master Mix, 1.55 µL H_2O, 0.9 µL of I and D allele-specific probes and 0.38 µL of *ACE* primer 111, 112, 113 (sequences below) were combined with 0.5 µL DNA solution (~23 ng DNA) per well for blood and saliva. For DNA derived from buccal cells, primer and probe volumes were identical but 0.05 µL H_2O and 2 µL DNA solution (~18 ng DNA) were used. Similarly, in the Northampton laboratory, *ACE* I/D was genotyped by combining 11 µL of Genotyping Master Mix, 2 µL of I and D probes, 2 µL of *ACE* primer 111, 112, 113 and 4 µL DNA solution (~40 ng DNA). Either a Chromo4 real-time system or a StepOnePlus™ was used. Briefly, there were 50 cycles of denaturation at 95°C for 15 s then annealing and extension at 57°C for 1 min. Initial analysis was

performed using Opticon Monitor 3.1 software or StepOnePlus™ software version 2.3. Again, there was 100% agreement within duplicates of all samples.

Primers and probes

For rs1815739 and rs4341, the appropriate TaqMan assay was used (Applied Biosystems). For the direct *ACE* I/D assay, three primers (150 nM each) and probes (VIC, 150 nM and FAM, 75 nM; Koch *et al.*, 2005) were used;

Primer ACE111: 5'-CCCATCCTTTCTCCCATTTCTC-3'

Primer ACE112: 5'-AGCTGGAATAAAATTGGCGAAAC-3'

Primer ACE113: 5'-CCTCCCAAAGTGCTGGGATTA-3'

I Allele specific probe (VIC-ACE100): VIC-5'AGGCGTGATACAGTCA-3'-MGB

D Allele specific probe (FAM-ACE100): FAM-5'TGCTGCCTATACAGTCA-3'-MGB

Positional groups

As detailed in Chapter *3.1.5*, to assess genotype and allele frequencies within the players, athletes were allocated in front five (props, hookers, locks), back row (flankers, number eights), half backs (scrum halves, fly halves), centres and back three (wings and full backs).

Data analysis

SPSS for Windows version 22 (SPSS Inc., Chicago, IL) software was used to conduct Pearson's Chi-square (χ^2) tests to compare genotype and allelic frequencies between athletes and controls, and between positional subgroups. For *ACTN3* and *ACE*, 26 and 16 tests, respectively, were subjected to Benjamini-Hochberg (BH; Benjamini & Hochberg, 1995) corrections to control false discovery rate and corrected probability values are reported. Where appropriate, OR was calculated to estimate effect size. Alpha was set at 0.05.

4.1.3 Results

All genotype data for athletes and controls were in Hardy-Weinberg equilibrium. Athletes were taller and heavier ($P < 0.0005$) but not older ($P = 0.871$) than controls.

There were no differences in genotype frequencies within the athlete or control groups according to nationality. For *ACE* I/D, there were no differences between all athletes and controls in genotype ($P = 0.83$), nor between playing subgroups (Table 6). Furthermore, for *ACTN3* R577X there were no genotype differences between controls and all athletes ($P = 0.33$). However, when considering playing position, the X allele was overrepresented in forwards (52.5%) compared to backs (37.8%; $P = 0.02$; OR = 1.49, 95%CI = 1.13-1.96, $P = 0.004$) and controls (42%; $P = 0.02$; OR = 1.25, 95%CI = 1.02-1.54, $P = 0.033$; Table 6 & Figure 19**A**). Similarly, there was a tendency ($P = 0.023$ before BH correction) of the XX genotype to be overrepresented in forwards (24.8%) compared to backs (15.7%; $P = 0.09$; OR = 1.77, 95%CI = 1.09-2.89, $P = 0.022$) and controls (18.3%; $P = 0.09$), with no difference between backs and controls ($P = 0.37$).

Table 6 Genotype and allele distribution of controls and athletes divided into positional subgroups for *ACE* and *ACTN3*, presented as genotype/allele counts followed by percentage in parentheses.

Genotype	RU athletes	Controls	Forwards	Front 5	Back row	Backs	Half Backs	Centres	Back three
ACE									
II	92 (21.5)	113 (19.8)	49 (20.0)	36 (22.1)	13 (15.9)	43 (23.6)	14 (20.3)	14 (31.1)	15 (22.1)
ID	214 (50.1)	286 (50.0)	129 (52.7)	86 (52.8)	43 (52.4)	85 (46.7)	33 (47.8)	17 (37.8)	35 (51.5)
DD	121 (28.3)	172 (30.2)	67 (27.3)	41 (25.2)	26 (31.7)	54 (29.7)	22 (31.9)	14 (31.1)	18 (26.5)
Total	427	572	245	163	82	182	69	45	68
I allele	398 (46.6)	512 (44.7)	227 (46.3)	158 (48.5)	69 (42.1)	171 (47.0)	61 (44.2)	45 (50.0)	65 (47.8)
D allele	456 (53.4)	630 (55.3)	263 (53.7)	168 (51.5)	95 (57.9)	193 (53.0)	77 (55.8)	45 (50.0)	71 (52.2)
ACTN3									
XX	90 (20.9)	130 (18.3)	61 (24.8)	39 (23.8)	22 (26.8)	29 (15.7)	12 (17.4)	11 (23.4)	*****6 (8.7)
RX	194 (45.0)	337 (47.5)	112 (45.5)	71 (43.3)	41 (50.0)	82 (44.3)	29 (42.0)	22 (46.8)	31 (44.9)
RR	147 (34.1)	[#]243 (34.2)	[#]73 (29.7)	54 (32.9)	19 (23.2)	74 (40.0)	28 (40.6)	14 (29.8)	32 (46.4)
Total	431	710	246	164	82	185	69	47	69
X allele	374 (43.4)	*****597 (42.0)	234 (47.6)	149 (45.4)	85 (51.8)	*****140 (37.8)	53 (38.4)	44 (46.8)	43 (31.2)
R allele	488 (56.6)	[#]823 (58.0)	258 (52.4)	179 (54.6)	79 (48.2)	230 (62.2)	85 (61.6)	50 (53.2)	*****95 (68.8)

RU, rugby union. *****Different from forwards. [#] Different from the Back three.

Interestingly, the 69 back three athletes (wings and fullbacks) included only six individuals (8.7%) of XX genotype which differed from the forwards (24.8%; P = 0.05; OR = 3.46, 95%CI = 1.43-8.34, P = 0.006) and tended to differ from the combined half backs and centres group (19.8%; P = 0.08; OR = 2.59, 95%CI = 1.00-6.74, P = 0.049). Likewise, the R allele distribution was greater in the back three (68.8%) than the controls (58.0%; P = 0.02; OR = 1.60, 95%CI = 1.09-2.33, P = 0.014), forwards (47.5%; P = 0.01; OR = 2.00, 95%CI = 1.34-2.99, P = 0.0007) and the other backs (58.2%; P = 0.05; OR = 1.59, 95%CI = 1.02-2.48, P = 0.042; (Figure 19**B**).

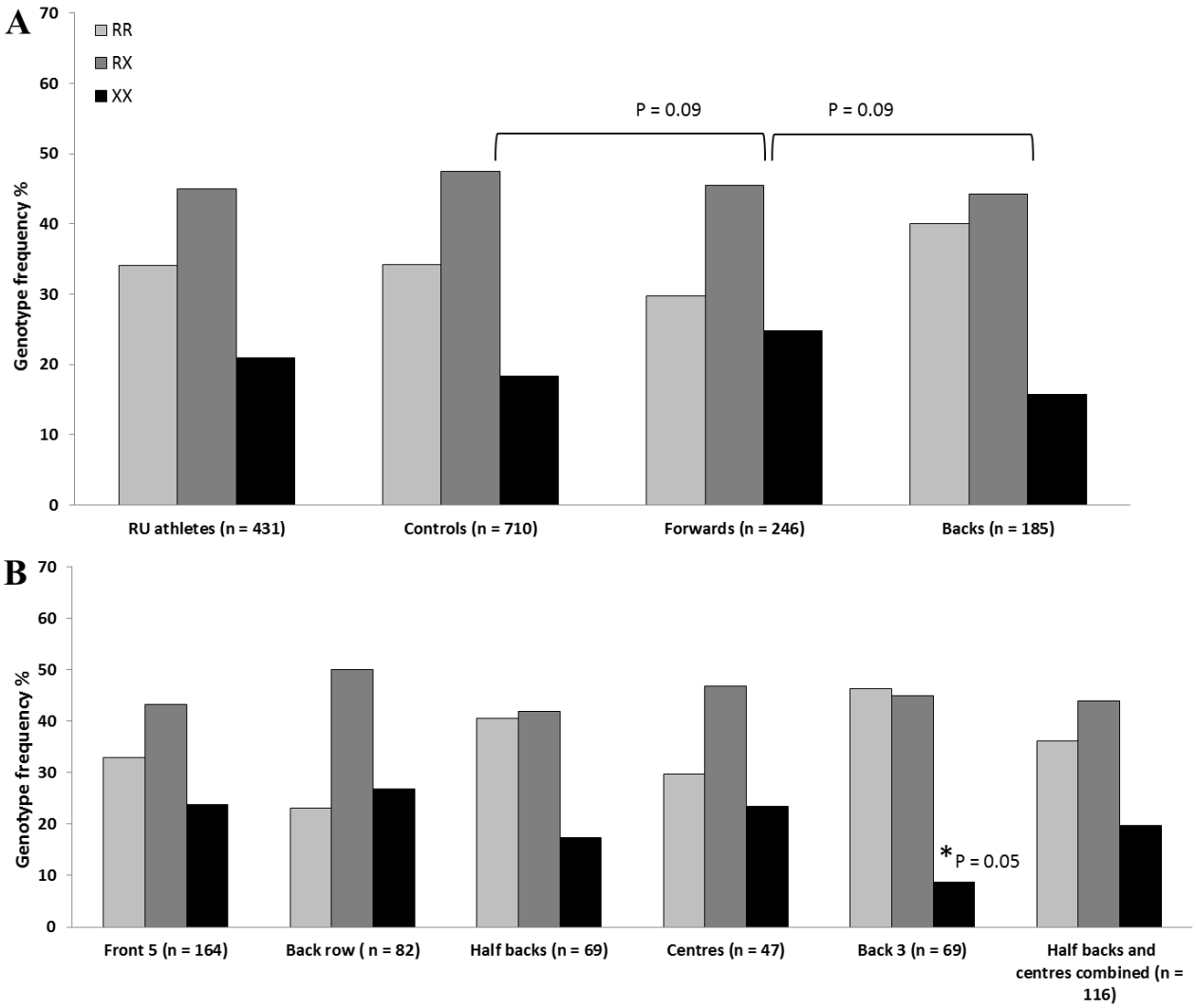


Figure 19 *ACTN3* genotype frequencies. **A**, Genotype frequency of athletes and controls, with athletes also divided into playing subgroup (forwards and backs). **B**, Genotype frequencies of RU athletes divided into positional groups with the addition of the 'half backs and centres combined' group. Statistical analysis between these positional groups was only performed to compare the back three with the half backs and centres combined. * Different from forwards.

4.1.4 Discussion

The present chapter is the first to show a genetic association with elite athlete status in rugby union. Associations for the *ACTN3* R577X polymorphism but not for *ACE* I/D were identified, thus rejecting the hypotheses regarding *ACE* I/D. Furthermore, no difference was observed for the *ACTN3* R577X genotype or allele distribution between all athletes and controls, when playing position was not considered, thus rejecting the hypothesis that differences would exist between non-athletes and all players as a single cohort. However, as hypothesized, in backs compared to forwards there was a lower proportion of XX genotype and X allele, which probably reflects the greater need for speed generation in backs and more sustained activity in forwards.

ACTN3 R577X

The most remarkable finding of the present study was the low frequency of the XX genotype among the back three athletes (8.7%), approaching although not as low as the frequency observed in elite sprinters (Yang *et al.*, 2003; Niemi & Majamaa, 2005). The XX genotype is present in ~18% of Caucasians (Table 6) and indicates absence of the α-actinin-3 protein (Beggs *et al.*, 1992; Mills *et al.*, 2001). Absence of α-actinin-3, a protein almost exclusively expressed in fast twitch skeletal muscle fibres, could hinder back three (wing and full back) sprint ability. R allele carriers have a greater proportion of type II and IIx fibres and larger relative surface area per IIx fibre than XX carriers (Vincent *et al.*, 2007; Ahmetov *et al.*, 2011; Broos *et al.*, 2012). Furthermore, Seto *et al* (2013) recently showed the likely mechanism for this genotype-phenotype association is via the calcineurin muscle fibre remodeling pathway. They found greater calcineurin activity (which induces slow myogenic programming and a shift towards oxidative phenotype) in α-actinin-3 knockout mice (KO) and humans (*ACTN3* 577XX genotype) due to preferential binding of α-actinin-2 (upregulated in the absence of α-actinin-3) to the fast fibre-specific calsarcin-2 (an inhibitor

of calcineurin). This could explain the advantage of R allele carriers over α-actinin-3 deficient XX individuals for high velocity contractions – particularly important for back three players. While backs and forwards previously showed similar fibre type proportions (Jardine *et al.*, 1988), these older data are arguably not relevant to modern rugby athletes, given their changed physical characteristics in recent years (Fuller *et al.*, 2013). Skeletal muscle fibre type proportions are unknown in contemporary elite rugby athletes who now compete in a more popular, fully professional sport and complete much higher training loads than previously. Recent *in vivo* data also shows that R allele carriers exhibit greater muscle volume and maximal power output (Gomez-Gallego *et al.*, 2009; Erskine *et al.*, 2014). While forwards show greater maximal power, backs are able to generate greater power relative to body mass ($W \cdot kg^{-1}$; Crewther *et al.*, 2012), which corresponds with the greater R allele frequency in the backs and especially the back three players. In fact, Broos et al. (2016) showed corresponding single fibre characteristics. Fibres of RR humans showed greater contractile velocity than XX individuals while exhibiting similar isometric force production. This suggests a strong rationale for the R allele advantage in elite sprinters (Yang et al., 2003; Niemi & Majamaa, 2005) and the back three athletes of the present chapter (Table 6 and Figure 19). This also supports the *ACTN3* R allele having a more relevant association with relative muscle power as opposed to absolute power (Kikuchi *et al.*, 2014b), due to preservation of force at high contractile velocities - a crucial relationship for sprinting performance (Miller *et al.*, 2012; Morin *et al.*, 2012) - and less so at reduced velocities (Broos *et al.*, 2016). As such, those rugby athletes where relative muscle power is an important quantity (Crewther *et al.*, 2012) would have a greater selective advantage with position of the R allele, as evidenced in the results of the present results (Table 6 and Figure 19). These data, plus evidence that type II fibres are larger and more powerful per unit volume than type I (Gilliver *et al.*, 2009), suggest the R allele would benefit back three rugby athletes for muscle power and fast fibre characteristics - which supports the present findings (Table 6 and Figure 19).

Arguably, the higher propensity for aerobic enzyme activity (porin, COX IV, hexokinase, citrate synthase, succinate dehydrogenase and β-hydroxyacyl CoA dehydrogenase; Seto *et al.*, 2011; Seto *et al.*, 2013) and greater force recovery after fatigue observed in α-actinin-3 deficient mice (Seto *et al.*, 2011), could indicate that XX genotype humans might have a greater capacity for recovery from fatiguing exercise - a trait which would benefit forwards with their more sustained match play intensity and necessity for quick recovery. The shorter rest periods for forwards compared with backs (work to rest ratios 1:7.4 and 1:21.8, respectively; Deutsch *et al.*, 2007) indicates that greater fatigue resistance would be particularly beneficial for forwards. Moreover, the greater calcineurin activity in XX homozygote humans and approximately threefold increase in calcineurin activity and distance run after endurance training in KO mice (Seto *et al.*, 2013), further support the notion that forwards would benefit from a greater fatigue resistance, especially with exposure to extensive training. These data are consistent with the observation that forwards exhibit higher XX genotype and lower R allele frequencies than backs and controls (Table 6). Additionally, these data could explain the present tendency for different allele frequencies between the forwards and backs ($P = 0.09$; $P = 0.023$ before BH correction; Figure 19). In that R allele carriers have a greater ability to achieve high velocity contractions – particularly important for back three players (Jones *et al.*, 2015) – and α-actinin-3 deficient XX individuals may have a greater capacity for recovery from repeated sprints – particularly important for the forwards (Deutsch *et al.*, 2007; Austin *et al.*, 2011a).

When considering many sports simultaneously, team sport athlete status showed no association with *ACTN3* R577X genotype (Eynon *et al.*, 2014). However, due to a relatively small number of athletes (205) with mixed status (56.6% elite) from a range of sports (ice hockey, handball, soccer, etc.), that is perhaps not surprising. While combining cohorts from different sports can boost sample size and theoretically increase statistical power, if an

association does not exist in all sports, or even in all athletes within a particular sport due to positional differences, one would be less likely to detect an association. The positional differences identified within the present study demonstrate the value of studying a large sample from a single sport and, in the absence of detailed physiological data (often difficult to obtain from large numbers of elite athletes), provides a viable alternative for future genetic research involving team sport athletes.

ACE I/D

The current study reports no difference between rugby athletes and controls or any positional subgroups for *ACE* I/D. This lack of association contrasts with a recent meta-analysis where the *ACE* I allele was associated with physical performance (Ma *et al.*, 2013).

However, Ma et al. also reported that specialized distance/endurance athletes showed the strongest association with the I allele (OR 1.35). Given the mixed metabolic nature of rugby, a comparable association in the present study was less likely. Furthermore, the importance of *ACE* I/D remains controversial in the literature, with no associations reported in other isolated team sports such as elite European soccer (Gineviciene *et al.*, 2014) and non-elite rugby athletes (Bell *et al.*, 2010). These prior data, in conjunction with the current findings in a larger study that also considers playing position, suggest that *ACE* I/D plays little role in performance of team sport athletes. *ACE* I/D genotype-athlete phenotype associations are more likely to exist in specialized endurance athletes (Puthucheary *et al.*, 2011).

Effect size and future applications

Odds ratios were calculated to estimate the likelihood that individuals with the advantageous genotype/allele become an elite rugby athlete in a specific position. The *ACTN3* XX genotype was almost twice (OR = 1.77) as common in forwards than backs, which suggests α-actinin-3 deficient individuals are more suited to forward play. Furthermore, forwards were over three times (OR = 3.46) more likely to be XX genotype than the back three athletes, while the remaining backs (centres and halves) were over twice as likely to show the α-actinin-3 deficient genotype than the back three (OR = 2.59). These data suggest the *ACTN3* R577X polymorphism shows potential to contribute to position-specific player profiling when combined with other genetic and physiological data in the future. In contrast, the *ACE* I/D polymorphism (OR ~1) does not show equivalent potential in rugby.

While the present cohort size is large compared to previous single sport genetic analyses, when the cohort was subdivided into playing position, the numbers were reduced so enlargement of the present cohort and replication would be welcome. Accordingly, the RugbyGene project continues to recruit elite rugby union players, so will steadily become better able to investigate genetic aspects of specific demands within rugby.

Chapter 5

Fat mass and obesity associated (FTO) gene influences elite rugby union playing position.

Data from this chapter are published as:

Heffernan, S.M, G. K. Stebbings, L. P. Kilduff, R. M. Erskine, S. H. Day, C. I. Morse, J. S. McPhee, C. J. Cook, B. Vance, W. J. Ribbans, S. M. Raleigh, C. Roberts, M.A. Bennett, G. Wang, M. Collins, Y. P. Pitsiladis & A. G. Williams. Fat mass and obesity associated (FTO) gene influences skeletal muscle phenotypes in non-resistance trained males and elite rugby playing position. *BMC genetics*, 18(1), 4.

5.1 Fat mass and obesity associated (*FTO*) gene influences elite rugby union playing position.

5.1.1 Introduction

Fat mass and obesity associated (*FTO*) is the most investigated gene in obesity and has complex molecular mechanisms that are yet to be fully elucidated. Recent genome-wide association studies (GWAS) have identified several common single nucleotide polymorphisms (SNP) in the human *FTO* gene associated with obesity, body mass index (BMI; Jacobsson *et al.*, 2012), cardiovascular disease and hypertension (Liu *et al.*, 2013; He *et al.*, 2014). These *FTO* SNPs, which are in strong linkage disequilibrium ($r^2 > 0.80$), are located in a cluster on the first intron of chromosome 16 and consequently exhibit similar obesity-related traits (Loos & Yeo, 2014). Thus, within different *FTO* variants, those alleles that have been positively associated with obesity-related phenotypes are referred to as risk alleles, while those negatively associated with such traits are referred to as protective alleles. Homozygotes for the minor risk allele consistently demonstrate greater BMI and body mass (3-10 kg) in comparison to protective allele carriers (Frayling *et al.*, 2007; Rauhio *et al.*, 2013). This greater body mass is likely to be adipose tissue (Andreasen *et al.*, 2008; Tanofsky-Kraff *et al.*, 2009; Liu *et al.*, 2010; Sonestedt *et al.*, 2011; Matsuo *et al.*, 2014), although there exist some suggestions of greater lean mass (LM) in addition to fat mass (Jess *et al.*, 2008; Sonestedt *et al.*, 2011).

Environmental lifestyle factors (diet and physical activity) have also been investigated for *FTO* gene-environment interactions. Risk allele carriers are more likely to choose a high fat diet than protective allele carriers (Tanofsky-Kraff *et al.*, 2009; Corella *et al.*, 2011; Phillips *et al.*, 2012). However, with administration of a high protein diet (25% energy intake) risk allele carriers demonstrated a greater reduction in body mass, fat mass and percentage body

fat (Zhang *et al.*, 2012), due to greater appetite suppression than in protective allele carriers (Huang *et al.*, 2014). Additionally, physically active risk allele carriers have a 30% reduction in likelihood of becoming obese and have 36% less body fat compared to inactive risk allele carrying individuals (Kilpeläinen *et al.*, 2011). Similarly, data from the HERITAGE Family Study showed that following 20 weeks of endurance training, protective allele homozygotes exhibited reductions in fat mass three times greater than risk allele carriers (Rankinen *et al.*, 2010). Interestingly, when comparing normal weight and obese individuals who participate in sport, no differences in *FTO* genotype were observed (P = 0.97), which was contrasted by those not participating (P = 0.02; Muc *et al.*, 2015). Considering the attenuation of *FTO*-associated obesity with environmental factors and the greater *FTO*-associated LM reported in obese populations (Jess *et al.*, 2008; Sonestedt *et al.*, 2011), investigating habitually trained elite athletes, in which body mass varies considerably across playing position, would be worthwhile.

To date, one study has considered *FTO* rs9939609 in athletic populations. Eynon et al. (2013b) investigated three European cohorts of power (n = 258; 58.3% elite) and endurance athletes (n = 266; 57.1% elite) from a variety of sporting disciplines - but identified no associations. This lack of association was likely due to the considerable differences in physiological demand between the various athletic disciplines included, plus further variability in the standard of athlete. The RugbyGene sample (Chapter 3) provides a unique opportunity to investigate *FTO* in individuals at the extreme upper end of physical fitness (100% elite athletes; Chapter *2*). Indeed, RU athlete positional divisions provide an ideal cohort to compare different anthropometric and physical quantises, while having some control over the environment of each athlete, as their training load and nutritional guidance are relatively similar. In terms of positional specific physiological differences that may be reflected in players' genetic variation, backs lower mass, lean mass, height and show lower

maximal strength and power measures compared to forwards. However, backs are faster, have a greater relative power output, cover more distance during games and have greater recovery than forwards (Table 1). Furthermore, the previous chapter has shown the ability of genetic research, in a single sport with player roles that differ distinctly, to reveal context-specific competitive advantages provided by particular alleles (Chapter *4*).

Therefore, the main aims of the present study were to investigate whether *FTO* rs9939609 genotype differed between elite rugby athletes and a control population, and/or between RU player positions. Based on prior data in obese populations, it was hypothesised that the rs9939609 risk (A) allele would be overrepresented in player positions typically requiring greater body and muscle mass while the protective (T) allele would be more common in positions requiring a lean phenotype.

5.1.2 Method

Participants

A total of 1089 individuals were recruited and gave written informed consent to participate in the present study. The total sample comprised elite Caucasian male rugby athletes (n = 450; height 1.85 (0.07) m, mass 101 (14) kg, age 29 (7) yr, BMI 29.4 (3.7) $kg \cdot m^{-2}$; mean (standard deviation (SD)) including 73% British, 16% South African, 7% Irish and 4% from other nationalities and non-athlete Caucasian control participants (60% male; n = 559; height 1.75 (0.10) m, mass 75 (13) kg, age 29 (16) yr, BMI 24.5 (3.6) $kg \cdot m^{-2}$) including 86% British, 12% South African, 1% Irish and 1% from other nationalities. Of the RU athletes, 52.7% had competed at an international level for a "High Performance Union" (Regulation 16, worldrugby.org). All data for the athlete group's international status were confirmed as of 1st June 2016.

Sample collection

Blood (n = 762 of all samples), saliva (n = 272) or buccal swab samples (n = 55) were obtained via the protocols detailed in Chapter *3.1.2*. Blood sample was drawn from a superficial forearm vein, into an EDTA tube and stored in 2 mL sterile tubes at -20˚C until processing. Saliva samples were collected into Oragene DNA OG-500 collection tubes according to the manufacturer's protocol and stored at room temperature until processing. Omni swab Sterile buccal were rubbed against the buccal mucosa of the cheek for approximately 30 s. Tips were ejected into sterile tubes and stored at -20˚C until processing.

DNA isolation

DNA isolation and genotyping were performed in the MMU, University of Glasgow, University of Cape Town (DNA isolation only) and University of Northampton laboratories. There are some differences between protocols summarized in Chapter *3.1.3*; however, there was 100% agreement among reference samples genotyped in the three genotyping centres, i.e. Glasgow, Northampton and MMU laboratories. The majority of samples were processed and genotyped in the MMU laboratory. Genotype calling was successful for both variants in all samples.

Genotyping

Genotyping in all three genotyping centres was performed on *FTO* (rs9939609). Briefly, in the Glasgow laboratory 10 μL Genotyping Master Mix (Applied Biosystems, Paisley, UK), 1 μL SNP-specific TaqMan assay (Applied Biosystems), 6 μL nuclease-free H_2O and 3 μL DNA solution (~9 ng DNA) were added per well. In the Northampton laboratory, genotyping was performed by combining 10 μL of Genotyping Master Mix, 8 μL H_2O, 1 μL assay mix with 1 μL of purified DNA (~10 ng). In both laboratories, PCR was performed using a

StepOnePlus™ real-time detector (Applied Biosystems). Briefly, after an initial 10 min at 95°C, 40 cycles of denaturation at 95°C for 15 s then annealing and extension at 60°C for 1 min were used. Genotyping calls were performed using StepOnePlus™ software version 2.3 (Applied Biosystems). At MMU, 5 μL Genotyping Master Mix, 4.3 μL H_2O, 0.5 μL assay mix and 0.2 μL of purified DNA (~9 ng) were used in each reaction for samples derived from blood and saliva. For DNA derived from buccal swabs, 5 μL Genotyping Master Mix was combined with 3.5 μL H_2O, 0.5 μL assay mix and 1 μL DNA solution (~9 ng DNA). Either a Chromo4 (Bio-Rad, Hertfordshire, UK) or StepOnePlus™ real-time PCR system was used. Briefly, after an initial 10 min at 95°C, 40 cycles of denaturation at 95°C for 15 s then annealing and extension at 60°C for 1 min were used. Genotyping calls were performed using Opticon Monitor software version 3.1 (Bio-Rad) or StepOnePlus™ software version 2.3. The Taqman assay included VIC and FAM dyes that indicated A and T alleles on the forward DNA strand, respectively. Thus, VIC/FAM were interpreted as: 5′-GTGAATTT[A/T] GTGATGCA-3′.

Positional groups

As detailed in Chapter *3.1.5*, to assess genotype and allele frequencies within the players, athletes were allocated in front five (props, hookers, locks), back row (flankers, number eights), half backs (scrum halves, fly halves), centres and back three (wings and full backs).

Data analysis

SPSS for Windows version 22 (SPSS Inc., Chicago, IL) software was used to conduct Pearson's Chi-square (χ^2) tests to compare genotype and allelic frequencies between athletes and controls, and between positional subgroups. Thirty tests were subjected to Benjamini-Hochberg (BH; Benjamini & Hochberg, 1995) corrections to control false discovery rate and

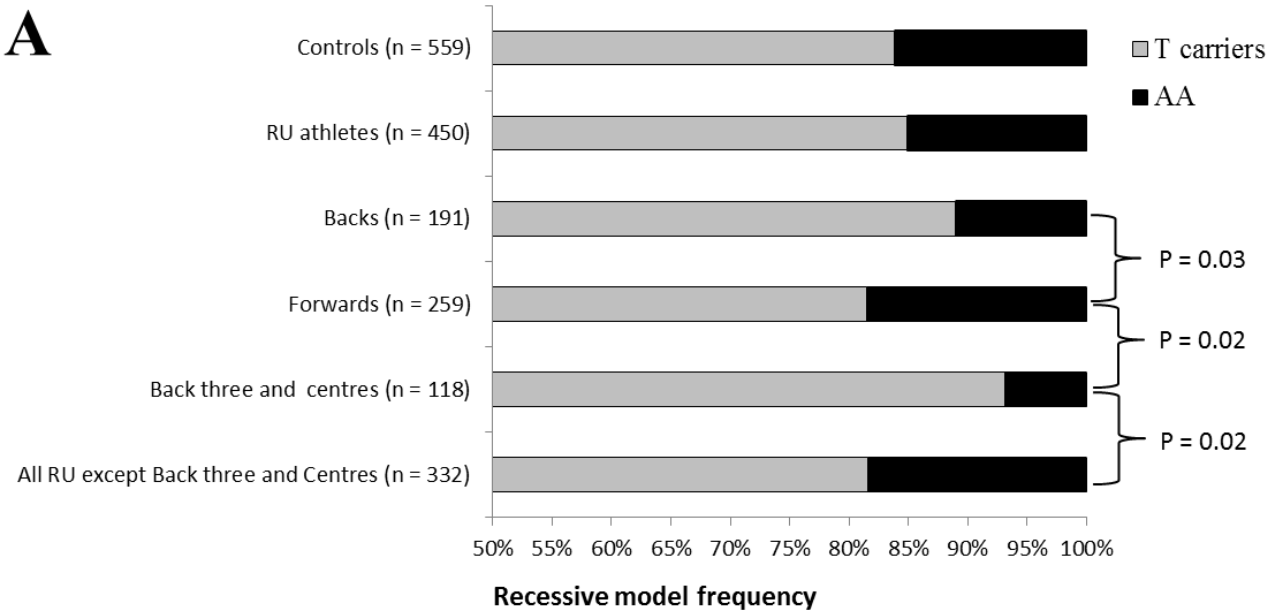
corrected probability values are reported. Where appropriate, OR was calculated to estimate effect size. Alpha was set at 0.05.

5.1.3 Results

Athletes were taller and heavier ($P < 0.05$) but not older ($P > 0.05$) than controls. There were no genotype frequency differences between RU athletes ($P = 0.16$) and controls (only additive models presented).

In terms of player position, backs had a greater frequency of T allele carriers than forwards ($P = 0.03$, Table 7, Figure 20) and showed greater odds of being T allele carriers than AA genotype (OR = 1.84, Table 8). When combined, the back three and centres group contained less AA homozygotes and more T allele carriers than controls ($P = 0.03$, $P = 0.02$, respectively; Figure 20A and Table 7). Additionally, controls had more than twice the odds of being AA than the back three and centres group, with greater odds of T allele carriers in the back three and centres than controls (Table 8).

Compared to forwards and all other RU athletes, TT genotype ($P = 0.03$; $P = 0.03$, respectively) and T allele carriers ($P = 0.02$; $P = 0.02$, respectively) were more common in the back three and centres group (Figure 20**A** and Table 7). Likewise, forwards and all other RU athletes had greater than three times the odds of being AA genotype than the back three and centres group, with greater odds of T allele carriers in the back three and centres group than forwards and all other RU athletes (Table 8). Furthermore, the back three and centres group showed a greater T allele frequency than both forwards and all other RU athletes (Figure 20**B**) and almost one and a half times greater odds of possessing the T allele (Table 8).



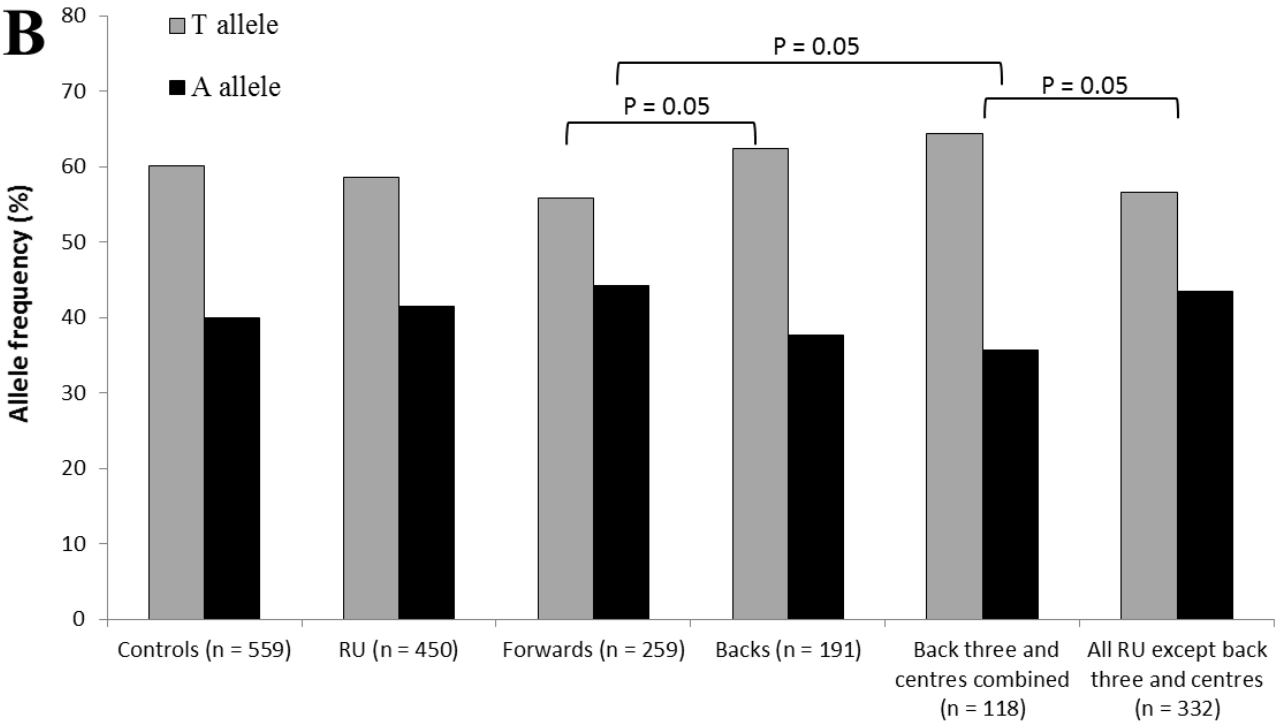


Figure 20 *FTO* genotype data of athletes and controls. **A** Recessive model. Grey bars are T allele carriers, black bars are AA genotypes. **B** Allele frequency for selected subgroups. Grey bars represent the T allele, black bars represent the A allele.

Table 7 Genotype and allele distribution of controls, RU athletes and divided into positional subgroups for *FTO*, presented as genotype/allele counts followed by percentage in parentheses.

Genotype/Allele	RU athletes	Controls	Forwards	Backs	Front 5	Back row	Half backs	Back three & centres
FTO								
AA	69 (15.3)	90 (16.1)	48 (18.5)	21 (11.0)	30 (17.0)	18 (21.7)	13 (17.8)	8 (6.8)*****
AT	235 (52.3)	266 (47.6)	133 (51.4)	102 (54.7)	94 (53.4)	39 (47.0)	34 (46.6)	68 (57.6)
TT	146 (32.4)	203 (36.3)	78 (30.1)	68 (34.3)	52 (29.6)	26 (31.3)	26 (35.6)	42 (35.6)‡
Total	450	559	259	191	176	83	73	118
A allele	375 (41.4)	446 (39.9)	229 (44.2)	144 (37.7)	154 (43.7)	75 (45.2)	60 (41.1)	84 (35.6)‡
T allele	527 (58.6)	672 (60.1)	289 (55.8)	238 (62.3)	198 (56.3)	91 (54.8)	86 (58.9)	152 (64.4)

RU, rugby union. *****Different from controls ($P < 0.04$). ‡Different from forwards ($P = 0.03$).

Table 8 Odds ratio statistics for RU athlete status by playing position for *FTO* genotype (TT/AA), allele (T/A) and recessive (T/AA) genetic models.

Positional Comparison	Genetic Model	Odds Ratio	95% Confidence Interval	P Value
Backs v Forwards	T/AA	1.84	1.06-3.19	0.029
Back three and centres v Controls	TT/AA	2.33	1.05-5.16	0.038
	T/AA	2.64	1.05-5.16	0.012
Back three and centres v Forwards	TT/AA	3.23	1.39-7.46	0.006
	T/AA	3.12	1.43-6.84	0.004
	T/A	1.44	1.04-1.97	0.026
Back three and centres v all other RU athletes	TT/AA	3.08	1.36-6.98	0.007
	T/AA	3.09	1.43-6.68	0.004
	T/A	1.37	1.01-1.86	0.045
Back three and centres v other backs	TT/AA	2.98	1.17-7.59	0.022
	T/AA	2.63	0.96-7.19	0.060

RU, rugby union.

5.1.4 Discussion

In agreement with the chapter hypothesis, the present data have shown that elite RU athlete playing positions more reliant on a lean phenotype for success (Smart *et al.*, 2013) possess a greater *FTO* rs9939609 T allele and TT genotype frequency than controls, while the A allele is more common in those positions where total body mass is more important (Sedeaud *et al.*, 2012; Figure 20, Table 7 and 8). The present findings disagree with Eynon et al. (2013b) who did not identify an association between *FTO* rs9939609, power (n = 258; 58.3% elite) and endurance (n = 266; 57.1% elite) athlete status. This lack of association was likely due to the considerable differences in physiological demand between the various athletic disciplines included together, plus further variability in the standard of athlete. It is likely that because of these methodological decisions, any possible association between *FTO* and athlete status might be filtered out and have produced false negative results. The present chapter differs in that the investigated cohort was a ubiquitous sample of 100% elite RU athletes from a single sport and importantly considered positional variation within RU. The present finding that the centre and back three group show a greater proportion of the T alleles suggest an advantage for these athletes in achieving elite RU status (Figure 20) and further show the importance positional variation within team sports. One possible biological mechanism underlying the present results may be the action of the iroquois homeobox 3 (IRX3) protein through its *FTO* genomic loci interaction.

Until recently, little was known about the molecular basis for *FTO* SNP associations with any reported phenotype measure, because there was no association between *FTO* SNPs and expression of the *FTO* protein (Wåhlén *et al.*, 2008; Grunnet *et al.*, 2009). However, *FTO* has recently been found to influence IRX3 protein expression, through evolutionarily conserved long-range chromatin looping. Individuals possessing the protective *FTO* genotype/allele (TT/T) display lower IRX3 expression than AA homozygotes (Smemo *et*

al., 2014). Furthermore, in contrast to IRX3 knockout (KO) mice, wild type mice exhibited similar *FTO* SNP risk (A) allele-associated phenotypes, such as greater BMI, body mass and body fat percentage (Smemo *et al.*, 2014), which reflects the present results of greater A allele frequency in the forwards playing position (Figure 20, Table 7 and 8) and suggests an advantage for the A allele and elite status in forwards - typically the those with the greatest mass (Table 1). Interestingly, IRX3 KO mice expended more energy, particularly at night, due to a greater browning of white fatty tissue (Smemo *et al.*, 2014) and recent findings show a link between brown fat and muscle developmental precursor Myf5 (Schulz *et al.*, 2011) which may provide a possible mechanism for the observation of greater *FTO* T allele carriers in the centre and back three cohort. Moreover, using a transgenic mouse model ($Rosa26^{Enr\text{-}Irx3}$) that disrupts IRX3 function whilst maintaining the genomic interaction between IRX3 and *FTO* (mimicking more accurately the human *in vivo* state than the aforementioned KO model), the authors showed retention of the KO model phenotype traits (Smemo *et al.*, 2014). These *FTO*-IRX3 protein interactions suggest a possible explanation for the results of the present chapter (Figure 20), in that T allele carriers (centres and back three players in the present chapter) may have a greater life-long predisposition to muscular development, however this has not yet been experimentally shown.

The precise mechanisms of action of IRX3 in mammalian physiology are not fully understood, however the primary role of IRX3 in embryonic development and future actions in motor neuron restriction is relevant to this discussion. During neuronal development, IRX3 expression plays a key role in N-tubulin development and initiation of neuronal programming. High levels of IRX3 protein promote early tissue development but not cell differentiation (Bellefroid *et al.*, 1998). It is possible that because the *FTO* T allele is associated with lower IRX3 expression, greater early differentiation might subsequently lead to greater muscular development. As such, for predeterminant neuronal cells to differentiate

into a progenitor motor neuron domain and subsequently motor neurons, it appears IRX3 expression must be repressed by the microRNA *mir-17-3p* in order for OLIG2 to regulate the development of ventral spinal motor neurons (Chen *et al.*, 2011). Thus, as the expression of OLIG2 increases, the yield of motor neurons increases in tandem (Lamas *et al.*, 2014). Considering *FTO* T allele carriers have a lower embryonic expression of IRX3, T allele carriers may have a predisposition for greater LM through enhanced life-long motor neuron availability via OLIG2 expression (diagrammatically shown in appendix 7) and therefore, may be at an advantage for certain forms of athletic ability and associated performance phenotypes (Table 7 and 8 ; Figure 20). This rationale and the present results are consistent with the 85% heritability of adult muscle neuronal function (Missitzi *et al.*, 2008).

Recent associations between *FTO* variants and IGF-1, specifically that serum IGF-1 levels were greater in T allele carriers (Rosskopf *et al.*, 2011), may provide a second mechanism to explain the observations of the present results. It is well known that IGF-1 is upregulated as a consequence of mechanical load/exercise and plays an important role in the cellular development of muscle hypertrophy (Sharples *et al.*, 2015). Hence, T allele carriers may experience upregulation of IGF-1 compared to AA genotype counterparts. These data provide a further potential basis for the observation that RU athletes who require a greater lean phenotype (Smart *et al.*, 2013; Smart *et al.*, 2014) and greater muscle power relative to body mass (Crewther *et al.*, 2012; center and back three group in the present chapter) show a greater frequency of the T allele than other playing positions (Table 7 and 8; Figure 20).

The present results observed a lower frequency of the AA genotype in back three and centre playing positions (OR = 2.53; Table 8), however there was no difference between the entire rugby cohort and controls. This demonstrates the importance of defining athletes very carefully when conducting such comparisons, as demonstrated previously regarding another

genetic variant *ACTN3* rs1815739 (Chapter 4). Global positioning system (GPS) data provide evidence for the relevance of the present finding regarding *FTO* genotype in elite athletes. Jones et al. (2015) showed that - at an elite competitive level - the back three and centre players express the greatest 'instantaneous and accumulative demands for exercise' (exertion index; EI) than all other athletes and spent more time at sprinting intensities.

These data suggest the relevance of the *FTO* rs9939609 T allele to athletic success for the backs playing positions and the A allele for the more massive forwards. Particularly, considering the T allele for athletic success in the backs, the possible molecular mechanism from *FTO* via IRX3 to OLIG2 resulting in greater lifelong motor neuron availability.

Chapter 6

Apolipoprotein E ε4 allele is not associated with elite rugby status but is present in 30% of athletes

This chapter is under review in Plos One as:

Heffernan, S. M., L. P. Kilduff, R. M. Erskine, S. H. Day, J, G. K. Stebbings, C. J. Cook, S. M. Raleigh, M. A. Bennett, G. Wang, M. Collins, Y. P. Pitsiladis & A. G. Williams. Apolipoprotein E ε4 allele is not associated with elite rugby status but is present in 30% of athletes. Submitted on 26/08/16.

6.1 Apolipoprotein E ε4 allele is not associated with elite rugby status but is present in 30% of athletes

6.1.1 Introduction

The *APOE* (apolipoprotein E) gene is located on chromosome 19, encodes apolipoprotein E-based peptide (ApoE) and is a candidate marker for risk and severity of mild-traumatic brain injury (mTBI) including sport-related concussion. ApoE is a 299 amino acid protein and has three common isoforms (Apo ε2, ε3, and ε4) which differ by two separate single amino acid changes (both cysteine/arginine). In humans, all three isoforms have a C-terminal domain that is largely responsible for lipid binding, whereas the N-terminus is comprised of a four α-helix motif that includes the receptor binding region (Laskowitz & Vitek, 2007). ApoE plays a pivotal role in cholesterol metabolism (Weisgraber, 1994; Huang & Mahley, 2014) and has been linked to neurobiological function (Teasdale *et al.*, 1997; Laskowitz *et al.*, 2010) with a particular susceptibility to late-onset and sporadic Alzheimer's disease via *APOE* gene polymorphisms (Busjahn *et al.*, 2009).

Carriers of the *APOE* ε4 allele have presented with reduced motor rehabilitation outcomes, poorer neurocognitive outcomes, increased cognitive impairments, amnesia and memory defects following traumatic brain injury (Lichtman *et al.*, 2000; Crawford *et al.*, 2002; Müller *et al.*, 2009; Noé *et al.*, 2010). Multiple meta-analyses have shown an increased risk of poor outcome greater than 6 months after mTBI (Zhou *et al.*, 2008; Zeng *et al.*, 2014; Li *et al.*, 2015) with one study suggesting that *APOE* ε4 was responsible for up to 64% of the 'hazardous influence' of TBI (Lawrence *et al.*, 2015). Of particular concern for athletic populations, where mTBI is generating increasing interest, Lawrence et al. (2015) reported that *APOE* ε4 was accountable for 38% of the 'hazardous influence' towards delaying recovery of mTBI (this analysis included, but was not limited to, concussion data). As

concussion is a form of mTBI and is classified as such (Roozenbeek *et al.*, 2013), the present chapter will refer to concussion as mTBI from here on.

There appears to be no association between *APOE* ε4 and self-reported history of sport-related mTBI (Terrell *et al.*, 2008; Tierney *et al.*, 2010) or mTBI diagnosed by a medical professional in a prospective study (Kristman *et al.*, 2008). This is not surprising, as large clinical studies show little *APOE* ε4 association with 'immediate' severity or morbidity but a more likely association with trajectory of recovery in subjects with severe TBI during rehabilitation (Noé *et al.*, 2010; Pruthi *et al.*, 2010; Ponsford *et al.*, 2011). Similarly, athletes carrying the *APOE* ε4 allele experience prolonged symptomatic responses to sport-related mTBI (Kutner *et al.*, 2000; Merritt & Arnett, 2016), which have recently been categorised as physical, cognitive, affective and sleep (Merritt & Arnett, 2016). In a recent study, Merritt and Arnett (2016) divided 42 college athletes who suffered an mTBI event into two groups: those possessing the ε4 allele (ε4+) and those not (ε4-). Consistently for all post-mTBI symptoms, ε4+ suffered more severely with a medium effect size for cognitive (Cohen's d = 0.60) and a large effect size for physical symptoms (d = 0.87) within 3 months post-mTBI.

Current mTBI incidence for elite rugby union (RU) players ranges between 4.6-8.9 per 1000 playing hours and has grown over time (Gardner *et al.*, 2014; Cross *et al.*, 2015; Fuller *et al.*, 2015a), with considerably higher rates in elite rugby league (RL), ranging between 14.8 - 28.3 per 1000 playing hours (Savage *et al.*, 2013; Gardner *et al.*, 2014; Gardner *et al.*, 2015b; Gardner *et al.*, 2016). Furthermore, in RU, data from the most recent Rugby World Cup (2015) indicates mTBI was the most common injury (14%; n = 24) and accounted for a total of 184 days absence (mean ~8 days each) from training and competition during and after that tournament (Fuller *et al.*, 2016). It is notable that players who returned to play in the same season following diagnosed mTBI have 60% greater chance of a subsequent time-

loss injury (not limited to head injury) than those that did not sustain mTBI (Cross *et al.*, 2015). As *APOE* ε4/ε4 genotype and ε4+ individuals are at greater risk of experiencing more severe symptoms after mTBI, it is probably more difficult for them to withstand repeated head impacts during many years of training and playing rugby without increasing incidence and severity of mTBI and other injuries. Consequently, those individuals would be forced to miss training, selection and competitive events important for their career progression, and thus might be at a disadvantage compared to ε4- individuals in terms of their ability to achieve success in elite competitive rugby.

Therefore, the aims of the present chapter were to establish the proportion of elite rugby athletes expected to be at higher risk of mTBI, due to carriage of the *APOE* ε4 allele and to investigate if *APOE* genotype differed between elite rugby athletes and a control population. Based on the published association of the *APOE* ε4 allele with poorer outcome following brain injury, it was hypothesised that the ε4/ε4 genotype and ε4+ allele would be underrepresented in elite rugby athletes compared to controls.

6.1.2 Method

Participants

A total of 926 individuals were recruited and gave written informed consent to participate in the present study. The sample comprised elite Caucasian male rugby athletes (n = 443; mean (standard deviation) height 1.85 (0.07) m, mass 101 (14) kg, age 29 (7) years) including 72% British, 16% South African, 7% Irish and 5% of other nationalities. Caucasian controls (61% male; n = 483; height 1.74 (0.10) m, mass 73 (13) kg, age 25 (11) years) included 98% British, 1% Irish and 1% of other nationalities. 51.9% of RU athletes had competed at international level for a "High Performance Union" (Regulation 16, worldrugby.org). All data for the athlete group's international status were confirmed as of 1st June 2016.

Sample collection

Blood (n = 704 of all samples), saliva (n = 252) or buccal swab samples (n = 50) were obtained via the protocols detailed in Chapter *3.1.2*. Blood sample was drawn from a superficial forearm vein, into an EDTA tube and stored in 2 mL sterile tubes at -20˚C until processing. Saliva samples were collected into Oragene DNA OG-500 collection tubes according to the manufacturer's protocol and stored at room temperature until processing. Omni swab Sterile buccal were rubbed against the buccal mucosa of the cheek for approximately 30 s. Tips were ejected into sterile tubes and stored at -20˚C until processing.

DNA isolation

DNA isolation and genotyping were performed in the MMU, University of Glasgow, University of Cape Town (DNA isolation only) and University of Northampton laboratories. There are some differences between protocols summarized in Chapter *3.1.3*; however, there was 100% agreement among reference samples genotyped in the three genotyping centres, i.e. Glasgow, Northampton and MMU laboratories. The majority of samples were processed and genotyped in the MMU laboratory. Genotype calling was successful for both variants in all samples.

Genotyping

Genotyping in all three genotyping centres was performed on *APOE* (rs429358 and rs7412). Briefly, in the Glasgow laboratory 10 μL Genotyping Master Mix (Applied Biosystems, Paisley, UK), 1 μL SNP-specific TaqMan assay (Applied Biosystems), 6 μL nuclease-free H_2O and 3 μL DNA solution (~9 ng DNA) were added per well. In the Northampton laboratory, genotyping was performed by combining 10 μL of Genotyping Master Mix, 8 μL H_2O, 1 μL assay mix with 1 μL of purified DNA (~10 ng). In both laboratories, PCR was

performed using a StepOnePlus™ real-time detector (Applied Biosystems). Briefly, after an initial 10 min at 95°C, 40 cycles of denaturation at 95°C for 15 s then annealing and extension at 60°C for 1 min were used. Genotyping calls were performed using StepOnePlus™ software version 2.3 (Applied Biosystems). At MMU, 5 μL Genotyping Master Mix, 4.3 μL H_2O, 0.5 μL assay mix and 0.2 μL of purified DNA (~9 ng) were used in each reaction for samples derived from blood and saliva. For DNA derived from buccal swabs, 5 μL Genotyping Master Mix was combined with 3.5 μL H_2O, 0.5 μL assay mix and 1 μL DNA solution (~9 ng DNA). Either a Chromo4 (Bio-Rad, Hertfordshire, UK) or StepOnePlus™ real-time PCR system was used. Briefly, after an initial 10 min at 95°C, 40 cycles of denaturation at 95°C for 15 s then annealing and extension at 60°C for 1 min were used. Genotyping calls were performed using Opticon Monitor software version 3.1 (Bio-Rad) or StepOnePlus™ software version 2.3. The TaqMan assay included VIC and FAM dyes that for rs429358 and rs7412 indicated C and T alleles on the forward DNA strand, respectively. Thus, for rs429358, VIC/FAM were interpreted as: 5′-AGGACGTG[C/T]GCGGCCGC-3′ and for rs7412 as: 5′-TGCAGAAG[C/T]GCCTGGCA-3′. The *APOE* gene ε2/ε3/ε4 haplotype was derived from rs429358 and rs7412 producing six possible genotypes (ε2/ε2, ε2/ε3, ε2/ε4, ε3/ε3, ε3/ε4, and ε4/ε4; Hixson & Vernier, 1990).

Data analysis

SPSS for Windows version 22 (SPSS Inc., Chicago, IL) software was used to conduct Pearson's Chi-square (χ^2) tests to compare genotype and allelic frequencies between athletes and controls, and between positional subgroups. Sixteen tests were subjected to Benjamini-Hochberg (BH; Benjamini & Hochberg, 1995) corrections to control false discovery rate and corrected probability values are reported. Where appropriate, OR was calculated to estimate effect size. Alpha was set at 0.05.

6.1.3 Results

Genotype calling was successful in all samples. There was 100% agreement among reference samples genotyped in the three genotyping centres, i.e. Glasgow, Northampton and MMU laboratories. Genotype frequencies for both rs429358 and rs7412 were in Hardy-Weinberg equilibrium for the athletes ($P > 0.27$) and control group ($P > 0.24$). Athletes were taller and heavier ($P < 0.05$) but not older ($P > 0.05$) than controls.

There were no differences in *APOE* genotype or ε4+ frequency when comparing all athletes ($P = 0.48$) with controls (additive values presented). Furthermore, no genotype frequency or ε4+ differences were observed between RU backs and forwards ($P = 0.89$, $P = 0.52$, respectively, Figure 21). However, despite RU athletes with international competitive experience having similar ε4/ε4 genotype frequency to controls (2.6% vs 2.3%), ε4/ε4 genotype frequency in those international athletes was lower ($P = 0.01$) than in athletes without international experience (4.7%, Table 9; OR = 2.02, 95%CI = 0.71-5.73, $P = 0.19$), with no association between ε4+ and international competitive experience.

Table 9 Genotype distribution of controls and RU athletes sub-divided by position and international status for *APOE*, presented as genotype counts followed by percentage in parentheses.

Genotype	Controls	RU Athletes	RU Internationals	RU Non-Internationals	RU Forwards	RU Backs
APOE						
ε2/ε2	1 (0.2)	1 (0.2)	0	1 (0.5)	1 (0.4)	0
ε2/ε3	59 (12.2)	53 (11.9)	18 (7.8)	35 (16.4)	29 (11.5)	24 (12.6)
ε2/ε4	13 (2.7)	7 (1.6)	1 (0.4)	6 (2.8)	5 (2.0)	2 (1.0)
ε3/ε3	288 (59.6)	259 (58.5)	142 (61.7)	117 (54.9)	145 (57.5)	114 (59.7)
ε3/ε4	111 (23.0)	107 (24.2)	63 (27.4)	44 (20.7)	63 (25.0)	44 (23.0)
ε4/ε4	11 (2.3)	16 (3.6)	6 (2.6)†	10 (4.7)	9 (3.6)	7 (3.7)
Total	483	443	230	213	252	191

RU, rugby union. †Different from RU non-international ($P = 0.01$).

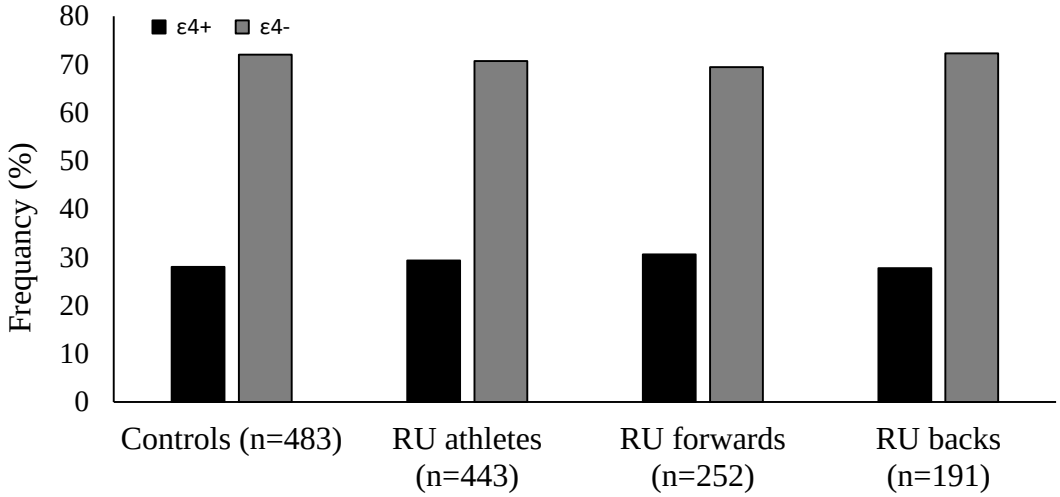


Figure 21 Presence or absence of ε4 allele in controls and athletes. Black bars represent ε4 carriers and grey bars represent those possessing no ε4 alleles.

6.1.4 Discussion

Here are the first data of *APOE* genotype and elite rugby athlete status. In contrast to the hypothesis, there were no differences in ε4/ε4 genotype or ε4 allele frequency between elite RU athletes and non-athlete controls (Figure 21). Furthermore, there were no differences in ε4/ε4 genotype or ε4 allele frequency between RU playing positions (Table 9).

In vivo, APOE ε4 has been associated with neurodegenerative cascade subsequent to TBI, the severity of axonal injury in mouse models (Sabo *et al.*, 2000; Hartman *et al.*, 2002; Bennett *et al.*, 2013) and more likely to show deposition of β-amyloid in brain tissue following head injury (Nicoll *et al.*, 1995; Teasdale *et al.*, 1997). In the intensive care setting, mice expressing the ApoE ε4 isoform have an increased systemic central nervous system inflammatory responses (Lynch *et al.*, 2003), while in humans the *APOE* ε4 polymorphism has been associated with increased systemic inflammatory responses (Moretti *et al.*, 2005). Considering these data with the observation of Merritt & Arnett (2016) and others (Lichtman *et al.*, 2000; Crawford *et al.*, 2002; Zhou *et al.*, 2008; Müller *et al.*, 2009; Noé *et al.*, 2010;

Roozenbeek *et al.*, 2013; Zeng *et al.*, 2014; Lawrence *et al.*, 2015; Li *et al.*, 2015) that ε4 allele carriers experience more severe cognitive and physical symptoms following mTBI, it was conceivable that the ε4 allele would be underrepresented in elite rugby athletes compared to controls. However, the present data do not support that hypothesis. As the present results show, the *APOE* ε4 allele appears to have no relationship with the ability of people to compete at the highest levels of competitive rugby (Figure 21), despite that being an environment of high risk of mTBI (Savage *et al.*, 2013; Gardner *et al.*, 2014; Cross *et al.*, 2015; Fuller *et al.*, 2015a; Gardner *et al.*, 2015a; Fuller *et al.*, 2016; Gardner *et al.*, 2016).

Notwithstanding previous reports (Kutner *et al.*, 2000; Merritt & Arnett, 2016) of ε4/ε4 genotype athletes or those carrying at least one ε4 allele being at increased risk of more severe symptoms following mTBI, that does not appear to inhibit their professional sporting careers. Three possibilities might explain this observation: (1) the published associations between the *APOE* ε4 allele and increased severity of mTBI in athletes are not true associations and will not be replicated in larger studies; (2) the increased risk associated with the ε4 allele is not large enough to influence career progression in professional rugby, as injured players are managed carefully, given appropriate recovery time and subsequently allowed to resume their careers; (3) the increased risk associated with *APOE* genotype is compensated, in terms of rugby career progression, by some advantageous role of the ε4 allele such as increased aerobic capacity or vitamin D availability (Hagberg *et al.*, 1999; Thompson *et al.*, 2004; Huebbe *et al.*, 2011; Raichlen & Alexander, 2014). Higher ε4/ε4 genotype frequency were observed in RU athletes without international experience compared to non-athletic controls (Table 9), which might indicate some advantage for competitive rugby performance. However, because this was not reflected in athletes with international experience, who showed no difference from controls, that observation should be treated with extra caution and emphasis most be placed on the data from the larger cohort

of all professional elite athletes regardless of international competitive experience (Figure 21). Furthermore, despite a previous investigation showing genetic variation related to muscle function and anthropometrics according to elite RU playing position (Chapters *4* and *5*), no association was identified between *APOE* genotype and RU playing position in the present chapter (Table 9).

It is noteworthy that there are considerable numbers of ε4/ε4 (3.3%) and ε4+ (29.8%) rugby athletes who may be at greater risk of cognitive and physical impairments following mTBI, compared to non-carriers (Lichtman *et al.*, 2000; Crawford *et al.*, 2002; Zhou *et al.*, 2008; Müller *et al.*, 2009; Noé *et al.*, 2010; Roozenbeek *et al.*, 2013; Zeng *et al.*, 2014; Lawrence *et al.*, 2015; Li *et al.*, 2015; Merritt & Arnett, 2016). World Rugby (the international governing body of rugby union) estimates there are 7.23 million rugby players worldwide of all competitive standards (*http://www.worldrugby.org/development/player-numbers*). Assuming similar allele frequencies to the Caucasians studied here amongst all players worldwide for the purpose of this discussion (i.e. ignoring geographic ancestry), over two million ε4+ rugby players may be at relatively greater risk of poorer outcome following mTBI than their ε4- counterparts. As Cross et al. (Cross *et al.*, 2015) has shown a 60% greater chance of time-loss injury in elite players who returned to competition in the same season as receiving a mTBI, correctly managing athletes that are at greater risk due to any predisposing factor such as the 30% who are ε4+, could become a valuable strategy for researchers, support scientists and medical staff in due course. Possession of the *APOE* ε4 allele is associated with differential immune responses related to nitric oxide synthesis via c-Jun N-terminal kinases (JNK) phosphorylation (Pocivavsek *et al.*, 2009a) and modulation of inflammatory processes involving TNF-α, IL -6 and IL-12 (Hidding *et al.*, 2002). Furthermore, the ε4 allele increases trauma-induced early apoptosis in neuronal/glial cell cultures (Chen *et al.*, 2015) while the ε3 isoform protects the blood brain barrier (BBB) by

controlling cyclophilin A (CypA) expression levels when compared to ε4 (Bell *et al.*, 2012b). These mechanisms produce a more hazardous cellular environment for ε4 carriers following injury. Further research to investigate if the previously identified association between *APOE* ε4 and mTBI severity (Lichtman *et al.*, 2000; Crawford *et al.*, 2002; Zhou *et al.*, 2008; Müller *et al.*, 2009; Noé *et al.*, 2010; Roozenbeek *et al.*, 2013; Zeng *et al.*, 2014; Lawrence *et al.*, 2015; Li *et al.*, 2015; Merritt & Arnett, 2016) persists in ε4 carriers amongst elite professional rugby players is therefore warranted.

Chapter 7

COL5A1 gene variants previously associated with reduced soft tissue injury risk are associated with elite athlete status in rugby union

This chapter is under review in BMC Genomics as:

Heffernan, S. M., L. P. Kilduff, R. M. Erskine, S. H. Day, J, G. K. Stebbings, C. J. Cook, S. M. Raleigh, M. A. Bennett, G. Wang, M. Collins, Y. P. Pitsiladis & A. G. Williams. *COL5A1* gene variants previously associated with reduced soft tissue injury risk are associated with elite athlete status in rugby union. Submitted on 16/09/16.

7.1 *COL5A1* gene variants previously associated with reduced soft tissue injury risk are associated with elite athlete status in rugby union

7.1.1 Introduction

Elite rugby athletes regularly experience high velocity collisions that lead to increasingly high injury occurrence rates that are likely to be an artefact of the increasing size and strength of the athletes (Cunniffe *et al.*, 2009; Sedeaud *et al.*, 2013; Owen *et al.*, 2014; Bradley *et al.*, 2015) and thus greater changes in momentum during player collisions, as well as during voluntary accelerations and decelerations. This has resulted in RU having one of the highest reported injury incidence rates in professional team sports (Brooks & Kemp, 2008). Meta-analyses have shown that for every 1000 hours, an elite RU athlete will experience approximately 81 injuries during match play and three during training, with the majority being ligament, tendon and muscle injuries of the lower limbs (Williams *et al.*, 2013). Indeed, in the most recent Rugby World Cup (2015) this rate of incidence was more than 90 injuries per 1000 h (Fuller *et al.*, 2016). Furthermore, injury incidence differs across RU playing position, with elite back row players showing the highest rate among forwards and centres the highest among backs (Fuller *et al.*, 2016). Therefore, investigating the molecular genetic components of these injuries, including in the context of playing positions that differ in terms of physiological characteristics (Smart *et al.*, 2013; Smart *et al.*, 2014), match play demands (Jones *et al.*, 2015) as well as genetically (Chapter 4 and 5), may progress understanding towards greater individualisation of match play exposure and training load and mode, in order to reduce injury risk.

The *COL5A1* gene, which encodes for a minor fibrillar collagen protein (Cappa *et al.*, 1995; Imamura *et al.*, 2000), is the most explored genetic locus in relation to tendon and ligament injuries. Collagen is the primary structural tissue protein of the extracellular matrix that in

animal models is suggested to regulate fibrogenesis (Chanut-Delalande *et al.*, 2004; Wenstrup *et al.*, 2006), with other non-human evidence suggesting this may be through altered fibril structure and diameter (Birk *et al.*, 1990). Furthermore, mutations in the human *COL5A1* gene disrupt collagen type V organisation and affect the assembly of other collagens in the extra cellular matrix (Zoppi *et al.*, 2004). Two amino acid components (collagen type V and type I fibrils) co-polymerise to form heterotypic fibres. The major collagen type V isoform comprises two α-1-(V) chains, encoded by the *COL5A1* gene, one α-2-(V) chain encoded by the *COL5A2* gene (Wenstrup *et al.*, 2004; Malfait *et al.*, 2010) and forms between 1-5% of total collagen content (McLaughlin *et al.*, 1989; Chanut-Delalande *et al.*, 2004). Mutations in the *COL5A1* gene have been identified in Ehlers-Danlos syndrome, a disease characterised by joint hypermobility, laxity and muscle hypotonia (Beighton *et al.*, 1998). This results in irregularly large collagen fibrils within connective tissue (Vogel *et al.*, 1979) and may be attributed to a reduced synthesis of collagen type V (Malfait & De Paepe, 2005; Sun *et al.*, 2011).

Two common *COL5A1* gene SNPs (rs12722 and rs3196378) located in the 3′ untranslated region (3′ UTR) on chromosome 9 have been associated with tendon (September *et al.*, 2009) and ligament (rs12722; Posthumus *et al.*, 2009a) pathology. Both rs12722 and rs3196378 were associated with tendinopathy in Australian Caucasians, but only the former in South African Caucasians (September *et al.*, 2009), which is interesting considering the SNPs are in linkage disequilibrium ($D' \geq 0.67$; September *et al.*, 2009; Laguette *et al.*, 2011). Accordingly, rs12722 has also been previously associated with flexibility (Collins *et al.*, 2009; Lim *et al.*, 2015), anterior cruciate ligament injury (Posthumus *et al.*, 2009b; Altinisik *et al.*, 2015; O'Connell *et al.*, 2015) and Achilles tendinopathy (Mokone *et al.*, 2006) but not patellar tendon dimensional or functional properties (Foster *et al.*, 2014). In these studies, the minor CC genotype was shown to be overrepresented in the respective asymptomatic

controls, suggesting a protective role of the C allele against injury. Considering the high frequencies of tendon and ligament inquiries in elite rugby (Williams *et al.*, 2013; King *et al.*, 2014; Fuller *et al.*, 2015a; Fuller *et al.*, 2016), assessing these specific genetic variants may be of use to help improve management of individual player injury risk.

Given the association of the two *COL5A1* gene variants with injury risk, it is probably more difficult for individuals carrying the alleles associated with greater risk to withstand exposure to the environment of competitive rugby without suffering from more frequent injuries. Consequently, those individuals would be forced to miss training, selection and competitive events important for their career progression. Thus, athletes carrying the C allele at both rs12722 and rs3196378 might be at an advantage in terms of their ability to achieve success in elite competitive rugby and at a disadvantage in terms of their shorter-term and longer-term musculoskeletal health. Therefore, the objective of the present study was to investigate if *COL5A1* rs12722 and rs3196378 genotype and allele frequencies differed between elite rugby athletes and a control population, and/or between playing positions. It was hypothesised that the *COL5A1* rs12722 and rs3196378 injury-protective C allele and CC genotype would be overrepresented in elite rugby athletes compared to controls.

7.1.2 Method

Participants

A total of 1020 individuals were recruited and gave written informed consent to participate in the present study. The sample comprised elite Caucasian male rugby athletes (n = 454; mean (standard deviation) height 1.85 (0.07) m, mass 101 (14) kg, age 29 (7) years) including 72% British, 16% South African, 7% Irish and 5% of other nationalities. Caucasian controls (62% male; n = 566; height 1.75 (0.10) m, mass 75 (13) kg, age 26 (11) years) included 86% British, 12% South African, 1% Irish and 1% of other nationalities recruited

mainly during 2012-2016. Of the RU athletes, 52.4% had competed at international level for a "High Performance Union" (Regulation 16, worldrugby.org). All data for the athlete group's international status were confirmed as of 1st June 2016.

Sample collection

Blood (n = 714 of all samples), saliva (n = 255) or buccal swab samples (n = 51) were obtained via the protocols detailed in Chapter *3.1.2*. Blood sample was drawn from a superficial forearm vein, into an EDTA tube and stored in 2 mL sterile tubes at -20˚C until processing. Saliva samples were collected into Oragene DNA OG-500 collection tubes according to the manufacturer's protocol and stored at room temperature until processing. Omni swab Sterile buccal were rubbed against the buccal mucosa of the cheek for approximately 30 s. Tips were ejected into sterile tubes and stored at -20˚C until processing.

DNA isolation

DNA isolation and genotyping were performed in the MMU, University of Glasgow, University of Cape Town (DNA isolation only) and University of Northampton laboratories. There are some differences between protocols summarized in Chapter *3.1.3*; however, there was 100% agreement among reference samples genotyped in the three genotyping centres, i.e. Glasgow, Northampton and MMU laboratories. The majority of samples were processed and genotyped in the MMU laboratory. Genotype calling was successful for both variants in all samples.

Genotyping

Genotyping in all three genotyping centres was performed on *COL5A1* rs12722 and rs3196378. Briefly, in the Glasgow laboratory 10 μL Genotyping Master Mix (Applied Biosystems, Paisley, UK), 1 μL SNP-specific TaqMan assay (Applied Biosystems), 6 μL nuclease-free H_2O and 3 μL DNA solution (~9 ng DNA) were added per well. In the Northampton laboratory, genotyping was performed by combining 10 μL of Genotyping Master Mix, 8 μL H_2O, 1 μL assay mix with 1 μL of purified DNA (~10 ng). In both laboratories, PCR was performed using a StepOnePlus™ real-time detector (Applied Biosystems). Briefly, after an initial 10 min at 95°C, 40 cycles of denaturation at 95°C for 15 s then annealing and extension at 60°C for 1 min were used. Genotyping calls were performed using StepOnePlus™ software version 2.3 (Applied Biosystems). At MMU, 5 μL Genotyping Master Mix, 4.3 μL H_2O, 0.5 μL assay mix and 0.2 μL of purified DNA (~9 ng) were used in each reaction for samples derived from blood and saliva. For DNA derived from buccal swabs, 5 μL Genotyping Master Mix was combined with 3.5 μL H_2O, 0.5 μL assay mix and 1 μL DNA solution (~9 ng DNA). Either a Chromo4 (Bio-Rad, Hertfordshire, UK) or StepOnePlus™ real-time PCR system was used. Briefly, after an initial 10 min at 95°C, 40 cycles of denaturation at 95°C for 15 s then annealing and extension at 60°C for 1 min were used. Genotyping calls were performed using Opticon Monitor software version 3.1 (Bio-Rad) or StepOnePlus™ software version 2.3. The TaqMan assay included VIC and FAM dyes that for rs12722 indicated C and T alleles on the forward DNA strand, respectively. Thus, VIC/FAM were interpreted as: 5′-CACACCCA[C/T]GCGCCCCG-3′. For rs3196378, VIC and FAM dyes indicated C and A alleles on the forward DNA strand, respectively and were interpreted as: 5′-CCCACCCC[A/C]GCCCTGGC-3′.

Data analysis

SPSS for Windows version 22 (SPSS Inc., Chicago, IL) software was used to conduct Pearson's Chi-square (χ^2) tests to compare genotype and allelic frequencies between athletes and controls, and between positional subgroups. Fifty tests were subjected to Benjamini-Hochberg (BH; Benjamini & Hochberg, 1995) corrections to control false discovery rate and corrected probability values are reported. Where appropriate, OR was calculated to estimate effect size. CubeX online software (*http://www.oege.org/software/cubex*) was used to determine linkage disequilibrium statistics (Gaunt *et al.*, 2007). Alpha was set at 0.05.

7.1.3 Results

Genotype calling was 100% successful for both polymorphisms in the athlete samples and for rs12722 in the control samples, with just 10 of the 566 control samples for rs3196378 unsuccessful. There was 100% agreement among reference samples genotyped in the three genotyping centres, i.e. Glasgow, Northampton and MMU laboratories. Genotype frequencies were in Hardy-Weinberg equilibrium for both rs12722 and rs3196378 in the control ($P \geq 0.09$) and athletes ($P \geq 0.78$). *COL5A1* rs12722 and rs3196378 were in tight linkage disequilibrium for both controls ($D' = 0.902$; $r^2 = 0.785$) and all athletes ($D' = 0.876$; $r^2 = 0.736$). Athletes were taller and heavier ($P < 0.05$) but not older ($P > 0.05$) than controls.

rs12722

The CC genotype and C allele were overrepresented in RU athletes (22.0% and 48.7%) compared to controls (13.6% and 41.3%, Table 10 and Figure 22, $P \leq 0.01$). Furthermore, the CC genotype (Table 10) and C allele (Figure 22) were overrepresented in the subgroups of RU forwards (22.3% and 48.7%) and backs (21.6% and 48.7%) compared to controls (13.6% and 41.3%). Additionally, of the RU subgroups, the back three and centres differed

from controls, showed the greatest C allele and CC genotype frequency (51.7% versus 41.3% and 24.8% versus 13.6%, respectively, Table 10 and Figure 22, P = 0.01) and had almost two and half times the odds possessing the C allele and CC genotype than controls (Table 11). There were no genotype or allele frequency differences any RU playing positions (Table 10).

Table 10 *COL5A1* rs12722 and rs3196378 genotype and allele distribution of controls and athletes separated into positional subgroups, presented as genotype/allele counts followed by percentage in parentheses.

Genotype	Controls	All athletes	RU athletes	Forwards	Backs	Back three & centres
rs12722						
TT	164 (29.0)	142 (26.2)	121 (26.7)	75 (28.8)	46 (23.7)	26 (21.5)
CT	325 (57.4)	279 (52.2)	233 (51.3)	127 (48.8)	106 (54.6)	65 (53.7)
CC	77 (13.6)	113 (21.2)*****	100 (22.0)*****	58 (22.3)*****	42 (21.6)*****	30 (24.8)*****
Total	566	534	454	260	194	121
T allele carriers	489 (86.4)	421 (79.9)*****	354 (78.0)*****	202 (77.7)*****	152 (78.4)*****	91 (75.2)*****
C allele carriers	402 (71.0)	392 (73.4)	327 (73.3)	185 (71.2)	148 (76.3)	95 (78.5)
rs3196378						
AA	183 (32.9)	142 (26.6)	121 (26.7)	74 (28.5)	47 (24.2)	26 (21.5)
CA	286 (51.4)	268 (50.2)	224 (49.3)	119 (45.8)	105 (54.1)	65 (53.7)
CC	87 (15.6)	124 (23.2)*****	109 (24.0)*****	67 (25.8)*****	42 (21.6)†	30 (24.8)*****
Total	556	534	454	260	194	121
A allele carriers	469 (84.4)	410 (76.8)*****	345 (76.0)*****	193 (74.2)*****	152 (78.4)*****	91 (75.3)*****
C allele carriers	373 (67.1)	392 (73.2)*****	333 (73.3)*****	186 (71.5)	147 (78.4)*****	95 (78.5)*****

RU, rugby union. *****Different from controls ($P \leq 0.03$)

rs3196378

The CC genotype, proportion of C allele carriers and C allele were overrepresented in RU athletes (24.0%, 73.3% and 47.7%) compared to controls (15.6%, 67.1% and 42.3%, Table 10 and Figure 22, $P \leq 0.02$). Furthermore, CC genotype, proportion of C allele carriers (Table 10) and C allele (Figure 22) were overrepresented in backs (21.6%, 75.8% and 48.7%) compared to controls (13.6%, 67.1% and 41.3%, $P \leq 0.02$). Forwards also had higher CC genotype and C allele frequencies (25.8% and 46.7%; Table 10 and Figure 22) and showed more than three and a half times the odds of being CC genotype than carrying an A allele, compared to controls (Table 11). For the back three and centres group, 24.8% were CC

genotype, 78.5% were C allele carriers and C allele frequency was 51.7% - all of which were greater than controls ($P \leq 0.02$; Table 10 and Figure 22, OR = 2.43, Table 11). However, there were no differences in rs3196378 allele or genotype frequencies between RU forwards and backs (Table 10).

Table 11 Odds ratio statistics for RU player status of *COL5A1* gene variants (rs12722 and rs3196378). Odds of the first genotype/allele in the first sample group.

Positional comparison	Genetic Model	Odds Ratio	95% Confidence Interval	P value
rs12722				
RU athletes v Controls	CC/TT	1.76	1.21-2.57	0.003
	C/T	1.24	1.04-1.48	0.015
Forwards v Controls	CC/TT	1.65	1.06-2.55	0.025
	CC/T	1.82	1.25-2.66	0.002
Backs v Controls	CC/TT	1.94	1.18-3.20	0.009
	CC/T	1.75	1.16-2.66	0.008
	C/T	1.31	1.04-1.65	0.023
Back three and centres v Controls	CC/TT	2.46	1.36-4.44	0.003
	CC/T	2.09	1.31-3.38	0.002
	C/T	1.46	1.10-1.92	0.008
rs3196378				
RU athletes v Controls	CC/AA	1.89	1.32-2.73	0.0006
	CC/A	1.70	1.24-2.33	0.0009
	C/A	1.35	1.13-1.61	0.0009
Forwards v Controls	CC/AA	1.90	1.25-2.89	0.003
	CC/A	3.45	2.21-5.40	<0.0001
	C/A	1.35	1.09-1.66	0.005
Backs v Controls	CC/AA	1.88	1.15-3.06	0.011
	C/A	1.35	1.07-1.70	0.011
Back three and centres v Controls	CC/AA	2.43	1.35-4.35	0.003
	CC/A	1.78	1.11-2.85	0.017
	C/A	1.52	1.15-2.01	0.003

RU, rugby union.

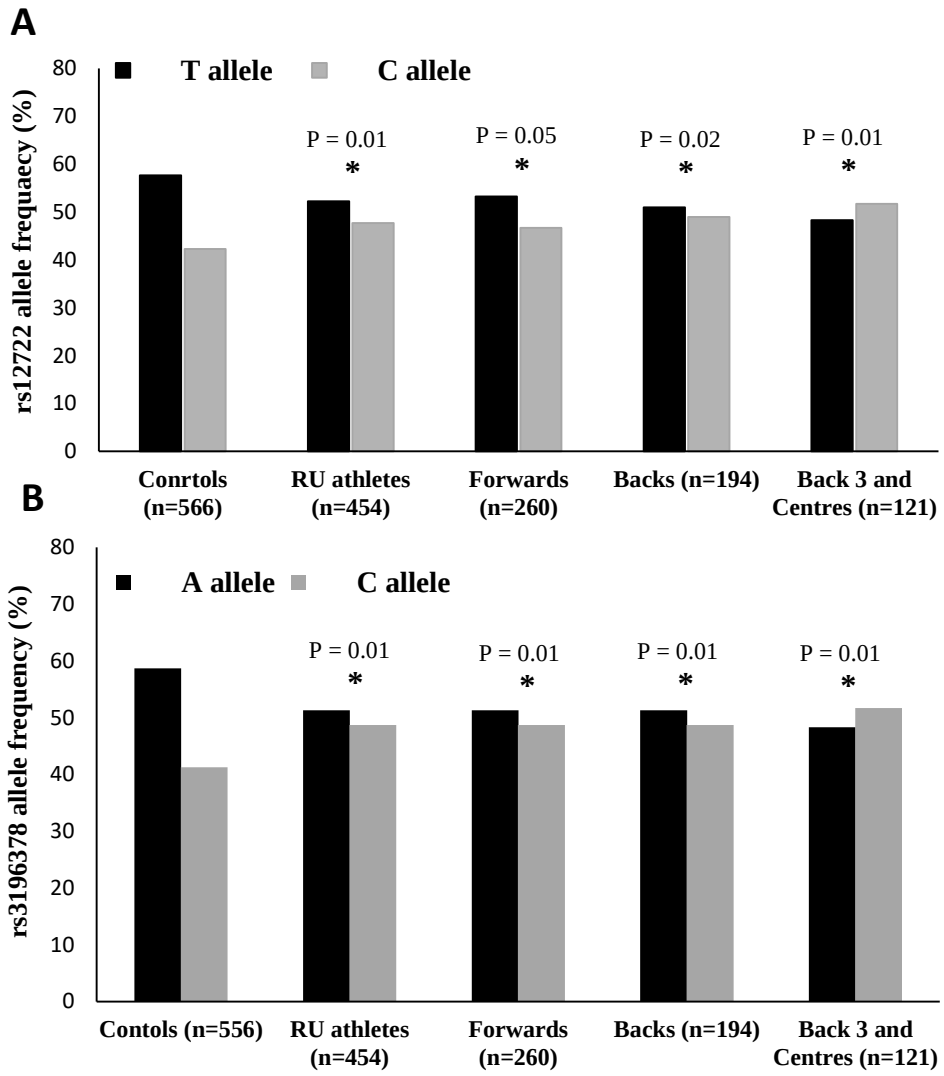


Figure 22 Allele frequency of *COL5A1* rs12722 (**A**) and rs3196378 (**B**) for controls and athlete groups. *Different from controls. RU, rugby union.

7.1.4 Discussion

The present observations are the first to identify associations between *COL5A1* rs12722 and rs3196378 polymorphisms and athlete status in a large cohort of elite RU athletes. As hypothesised, the injury-protective C allele and CC genotype, of both SNPs (September *et al.*, 2009), were over-represented in elite RU athletes, compared to controls (Table 10 and Figure 22). This association persisted across playing position, with the C allele being overrepresented in RU forwards and backs including the back three and centres group, compared to controls (Table 11 and Figure 22).

The results provide an insight into the potential injury susceptibility of some elite rugby athletes. September et al. (2009) identified a higher frequency of the CC genotype in asymptomatic controls for both rs12722 and rs3196378 compared to tendinopathy patients (September *et al.*, 2009; Laguette *et al.*, 2011). Moreover, the rs12722 T allele has been associated with anterior cruciate ligament injury (Posthumus *et al.*, 2009b; Bell *et al.*, 2012a; Altinisik *et al.*, 2015; O'Connell *et al.*, 2015) and Achilles tendinopathy (Mokone *et al.*, 2006; September *et al.*, 2009) with the C allele again identified as protective in these studies. Greater joint laxity has almost a 3-fold increase in risk of knee ligament rupture (Uhorchak *et al.*, 2003) and greater joint laxity has recently been associated with the rs12722 T allele in non-white females (Bell *et al.*, 2012a). These data suggest that the C allele of both rs12722 and rs3196378 may be beneficial in protecting against tendon and ligament injuries. This is reflected in the present chapter showing greater C allele frequency in elite RU athletes compared to controls (Table 10 and 11, Figure 22). Based on these data, it is proposed that when exposed to the high-risk environment of RU during training and especially during competitive matches, *ceteris paribus*, carriage of the C alleles at those two SNPs provides both a shorter-term and longer-term advantage to rugby athletes in the form of reduced injury risk. Athletes with fewer and/or less severe injuries, all else being equal, will miss fewer matches, training and selection events and thus be more likely to progress towards elite status in their athletic careers compared to their peers.

The rs12722 CC genotype has also been related to have a lower incidence of exercise-associated muscle cramping (EAMC) in Caucasian ironman and ultra-marathon athletes (O'Connell *et al.*, 2013). The authors hypothesised that this was due to similar mechanisms of reduced tendon and injury susceptibility, in that rs12722 alters soft tissue structural and mechanical properties (tissue thickening). However, current evidence does not support an association of rs12722 with structural and mechanical properties (Foster *et al.*,

2014).Regardless, this suggests that in addition to protection from tendon and ligament injury, the greater frequencies of the C allele in elite RU athletes (Table 10 and Figure 22) may be protective against muscle cramping. Indeed, recent evidence from elite rugby league athletes shows that over 70% of athletes experience EAMC per season and that history of cramping is the strongest predictor of future EAMC (Summers *et al.*, 2014). Which may be supported, to a lesser extent, in elite RU (23%; categorised in a subgroup with muscle rupture, tear and strain; Fuller *et al.*, 2016). In contrast, the TT genotype has been associated with greater endurance running ability of Caucasian ironman triathletes (TT = 294.2 min, CC = 307.4 min; Posthumus *et al.*, 2011). However, recent data shows no association of rs12722 with either running economy or $V\dot{O}_{2max}$ (Bertuzzi *et al.*, 2014). While endurance capacity is of value in elite rugby, the predominant focus of player selection and training programs is towards power, speed and strength - i.e. short-term, anaerobic performance (with notable differences between playing positions; Smart *et al.*, 2013; Smart *et al.*, 2014).

Limited data exist regarding *COL5A1* genetic variation and team sport athletes. In a study of 73 soccer athletes, including some elite players, no rs12722 TT genotype individuals were identified (a potentially interesting observation but difficult to interpret because of the varied geographic ancestry of the athletes), but there was a tendency for more severe muscle injuries in the TC genotype group (P = 0.08), compared to CC (Pruna *et al.*, 2013). Here, consistent with those observations, the data show an overrepresentation of the protective C allele and CC genotype of both rs12722 and rs3196378 in elite rugby athletes (with no differences between playing positions).

Some possible mechanisms have been proposed to explain the association of *COL5A1* gene variants and soft tissue injury (Laguette *et al.*, 2011; Abrahams *et al.*, 2013). Laguett et al. (2011) have shown that the *COL5A1* 3′ UTR – where both rs12722 and rs3196378 are

situated - affects miRNA stability. For both SNPs, the alleles associated with greater soft-tissue injury risk were associated with greater Hsa-miR-608 stability, which in turn may alter the Col5α1 protein secondary structure - proposed to play a role in type V collagen production (Abrahams *et al.*, 2013). This would suggest that C/T allele differences at rs12722 may alter the co-polymerisation of collagen type V and type I fibrils. However, to date, this has not been demonstrated experimentally and exactly how this may translate into functional properties is currently unknown. Nevertheless, it appears that the C allele and CC genotype of rs12722 and rs3196378 may be beneficial for rugby athletes to achieve elite status, probably through greater resistance to soft tissue injury. Interestingly, while most relevant investigations have focussed on rs12722, the present data show, in a large cohort (total n = 1090), that strong linkage disequilibrium exists in both controls ($D' = 0.902$; $r^2 = 0.785$) and athletes ($D' = 0.876$; $r^2 = 0.736$) between rs12722 and rs3196378. As such, it is likely that the molecular associations of rs12722 with tendon and ligament injuries would be similar for rs3196378. It is possible that combining genetic data from multiple gene variants associated with injury susceptibility, such as those presented here, with other indicators of injury risk and recovery during rehabilitation could be used to better manage the prevention of and recovery from elite player injury in the future.

Chapter 8

Polygenic profile of elite rugby union athletes

8.1 Polygenic profile of elite rugby union athletes

8.1.1 Introduction

Elite athletic performance is highly heritable (66%; De Moor *et al.*, 2007) and consists of complex physiological traits. The molecular genetic components of these traits have been receiving increasing interest from the scientific community in recent years (Ahmetov & Fedotovskaya, 2012; Ahmetov *et al.*, 2016; Pitsiladis *et al.*, 2016; Tanaka *et al.*, 2016) - *Chapter 2.2*. In fact, more than 155 genetic markers have been identified in association with athletic performance and confirmation of these results has been building with 31 publications showing positive associations in at least 2 studies (Ahmetov *et al.*, 2016). Because each individual genetic variant carries only a small contribution to the overall heritability of elite athletic status, using statistical models to combine these individual influences are required (Pitsiladis *et al.*, 2013), have been developed (Williams & Folland, 2008) and have recently been expanded upon (Bouchard *et al.*, 2011; Thomaes *et al.*, 2011; Thomaes *et al.*, 2013; Massidda *et al.*, 2014b).

Williams and Folland (2008) developed the total genotype score (TGS) as a way to represent the simultaneous influence of multiple genetic markers as a simple value, that is intuitively understandable. One of the main aims of the model is to combine the small accumulative effects of many SNPs, that may have been statistical insignificant due to statistical power via low sample size (a current limitation identified in chapter *2.2.4; appendix 5*), to give a better understanding of the polygenic nature of elite athletic traits. According to the TGS, all scores lie between 0-100 and a higher score indicates a greater genetic suitability for a given phenotype such as muscle mass, sprinting speed, reduced injury risk, greater cognitive ability, elite athlete status – depending on the genetic variants that are used to calculate the specific TGS in each case. Thus, multiple TGS, each directed towards a given phenotype,

can be calculated for each individual. Subsequently, both hypothetical (Williams & Folland, 2008; Hughes *et al.*, 2011) and experimental (Gómez-Gallego *et al.*, 2010; Ruiz *et al.*, 2010; Santiago *et al.*, 2010; Eynon *et al.*, 2011b; Chiu *et al.*, 2012; Drozdovska *et al.*, 2013; Massidda *et al.*, 2014a; Ahmetov *et al.*, 2015; Ben-Zaken *et al.*, 2015; Grealy *et al.*, 2015; Miyamoto-Mikami *et al.*, 2016) studies have used this method in attempts to elucidate the 'optimal polygenetic profile' for a given cohort with some limited success.

To date, TGS has been considered in only two team sport investigations of elite soccer athletes, one concerned with phenotypic measures of lower body power (Massidda *et al.*, 2014a) and the other with positional variation and athlete status (Egorova *et al.*, 2014). Massidda et al. showed no difference between TGS in the whole athlete cohort (n = 90) and controls, however did identify a relationship between TGS (of six SNPs) and vertical jump performance, using the Williams and Folland method (Williams & Folland, 2008; Massidda *et al.*, 2014b). In addition to the TGS, Massidda et al. developed a genotype score (GS) that weighted (W) each SNP based on their phenotypic impact resulting in three SNPs being excluded from the WGS and showed a greater explanation of the variance in vertical jump performance than non-weighted TGS. Weighting each SNP based on the relevance to a given phenotype is an important development, however does not currently apply to investigations of athlete status due to the categorical nature of the data. More applicable to the present thesis, Egorova et al. firstly performed a case-control study of eight SNPs in 246 mixed level athletes (n = 51 elite), where four SNPs were identified as significant in the whole group. These four SNPs were then included into the Williams and Folland TGS algorithm to compare athletic level and positional difference to controls in a pseudo-data drive approach. Strictly speaking, the Williams and Folland method was developed to include both significant and non-significant results to ensure identification of any possible association. Nevertheless, they found that at all competitive levels (elite, sub-elite and non-elite) athletes had greater TGS than controls and of the four positional groups (comprising of elite (n = 51)

and sub-elite (n = 83), goalkeepers, attackers, defenders and midfielders), goalkeepers and midfielders were independently different from controls (Egorova *et al.*, 2014), suggesting that TGS models may differ between team sport playing position and need to be considered within the present chapter. As shown in chapters *4* and 5, considerable genetic differences exist within elite rugby playing position and an exploratory polygenic investigation of the genetic variants studied in the present thesis is the logical next step. Indeed, because of the uniquely identified genetic differences between RU backs and forwards (chapters *4* and 5) assigning the optimal genotype score using evidence from the existing literature would introduce considerable errors to the model because the 'actual' (i.e. results from chapters *4* and 5) data would not be represented in the model. However, assigning the genotype score based on these uniquely identified genetic differences (data driven approach) and following the inclusion of all studied variants, as originally intended by the Williams and Folland method (Williams & Folland, 2008), would significantly improve the accuracy of the model.

Therefore, utilising the data driven approach to the Williams and Folland TGS algorithm, the aims of the present chapter were firstly to assess the polygenic profile, of the gene variants examined in the first four experimental chapters (chapters 4-7), for all RU athletes, forwards and backs, compared to controls. Secondly, because of the inter-positional genetic variation that has been identified (chapters *4* and 5), the present chapter will explore and compare the polygenic difference between backs and forwards.

8.1.2 Method

Participants

A total of 881 individuals were recruited and gave written informed consent to participate in the present thesis. The sample comprised elite Caucasian male rugby athletes (n = 422; mean (standard deviation) height 1.85 (0.07) m, mass 101 (14) kg, age 29 (7) years) including 72%

British, 16% South African, 7% Irish and 5% of other nationalities. Caucasian controls (60% male; n = 459; height 1.74 (0.10) m, mass 73 (13) kg, age 25 (11) years) included 98% British, 1% Irish and 1% of other nationalities. Of the RU athletes, 51.9% had competed at international level for a "High Performance Union" (Regulation 16, worldrugby.org). All data for the athlete group's international status were confirmed as of 1st June 2016.

Sample collection

Blood (n = 617 of all samples), saliva (n = 220) or buccal swab samples (n = 44) were obtained via the protocols detailed in Chapter *3.1.2*. Blood sample was drawn from a superficial forearm vein, into an EDTA tube and stored in 2 mL sterile tubes at -20˚C until processing. Saliva samples were collected into Oragene DNA OG-500 collection tubes according to the manufacturer's protocol and stored at room temperature until processing. Omni swab Sterile buccal were rubbed against the buccal mucosa of the cheek for approximately 30 s. Tips were ejected into sterile tubes and stored at -20˚C until processing.

DNA isolation

DNA isolation and genotyping were performed in the MMU, University of Glasgow, University of Cape Town (DNA isolation only) and University of Northampton laboratories. There are some differences between protocols summarized in Chapter *3*; however, there was 100% agreement among reference samples genotyped in the three genotyping centres, i.e. Glasgow, Northampton and MMU laboratories. The majority of samples were processed and genotyped in the MMU laboratory.

Genotyping

Genotyping in all three genotyping centres was performed on *APOE* (rs429358 and rs7412), *ACTN3* (rs1815739) and an *ACE I/D* or *ACE* tag SNP (rs4341), *COL5A1 (*rs12722 and rs3196378) and *FTO* (rs9939609). Briefly, in the Glasgow laboratory 10 *μL* Genotyping Master Mix (Applied Biosystems, Paisley, UK), 1 *μL* SNP-specific TaqMan assay (Applied Biosystems), 6 *μL* nuclease-free H_2O and 3 *μL* DNA solution (~9 ng DNA) were added per well. In the Northampton laboratory, genotyping was performed by combining 10 *μL* of Genotyping Master Mix, 8 *μL* H_2O, 1 *μL* assay mix with 1 *μL* of purified DNA (~10 ng). In both laboratories, PCR was performed using a StepOnePlus™ real-time detector (Applied Biosystems). Briefly, after an initial 10 min at 95°C, 40 cycles of denaturation at 95°C for 15 s then annealing and extension at 60°C for 1 min were used. Genotyping calls were performed using StepOnePlus™ software version 2.3 (Applied Biosystems). At MMU, 5 *μL* Genotyping Master Mix, 4.3 *μL* H_2O, 0.5 *μL* assay mix and 0.2 *μL* of purified DNA (~9 ng) were used in each reaction for samples derived from blood and saliva. For DNA derived from buccal swabs, 5 *μL* Genotyping Master Mix was combined with 3.5 *μL* H_2O, 0.5 *μL* assay mix and 1 *μL* DNA solution (~9 ng DNA). Either a Chromo4 (Bio-Rad, Hertfordshire, UK) or StepOnePlus™ real-time PCR system was used. Briefly, after an initial 10 min at 95°C, 40 cycles of denaturation at 95°C for 15 s then annealing and extension at 60°C for 1 min were used. Genotyping calls were performed using Opticon Monitor software version 3.1 (Bio-Rad) or StepOnePlus™ software version 2.3. The Taqman assay included VIC and FAM dyes, interpretation of the specific sequence for each SNP are presented in chapters *4*, *5*, *6* and *7*.

Positional groups

As detailed in Chapter *3.1.5*, to assess genotype and allele frequencies within the players, athletes were allocated to subgroups; forwards (props, hookers, locks, flankers, number eights) and backs (scrum halves, fly halves, centres, wings, full backs).

TGS calculations

To determine TGS', each genotype was allocated a 'genotype score' (GS) of 0, 1 or 2. The allocation was based on the assumption that allele effects were co-dominant, and homozygotes deemed to have the favourable genotype for the phenotypes of interest (athlete/positional status) were allocated a GS of 2, heterozygotes scored 1 and the non-favourable homozygotes scored 0. For *APOE*, the combination of rs429358 and rs7412 generates the ε4 genotype, in the context of an athletic population, the presence (ε4+) or absence (ε4-) of the ε4 allele have been identified as the best genetic model (for details see chapter *6*). As such, ε4+ were given the value 0 and ε4- were allocated 2. There was no *APOE* group allocated 1. Combination of each GS and transformation of the total score allowed the combined influence of all 6 gene variants on athlete and positional status to be quantified (Williams & Folland, 2008) using the following equation (equation one).

Equation one: $TGS = (100/12) * (GS_{APOE\text{-}\varepsilon4} + GS_{ACE} + GS_{ACTN3} + GS_{COL5A1\text{-}rs12722} + GS_{COL5A1\text{-}rs3196378} + GS_{FTO})$.

By following the experimental data (data driven approach), to ensure accuracy and specificity of the TGS model to RU athletes, presented in the previous chapters (chapter *4*, *5*, *6* and *7*) and including all investigated genotypes, regardless of a non-significant

association (Williams & Folland, 2008), decisions of the optimal GS for each genotype generated three TGS models. (1) Based on the observed data for all RU athletes (GS_{RU}; Table 12), (2) based on the observed data for the backs playing position (GS_B; Table 14) and (3) based on the observed data for the forwards playing position (GS_F; Table 13). Each model was calculated for both the athlete group and the control group and as such, comparisons between athletes (including backs and forwards) and controls were made from the same model. Additionally, in an attempt to normalise the TGS for positional specificity, a TGS_B/TGS_F ratio was calculated and compared between backs and forwards using receiver operating characteristic (ROC) curves, similar to that described in an alternative athletic context (Ben-Zaken *et al.*, 2015).

Table 12 Genotype score details for optimal **RU athlete** (GS_{RU}) status based on evidence from the thesis observations (genotype score chosen from the data contained within the experimental chapters, 4-7).

Gene variant	Genotype score	Rationale	Evidence
ACE (I/D)	0 = DD, 1 = ID & 2 = II	I allele associated with athletic performance, particularly endurance performance and slightly greater proportion in all elite rugby athletes.	Chapter 2.3.1 Chapter 4
ACTN3 rs1815739	0 = XX, 1 = RX & 2 = RR	R allele associated with sprint/power performance. X allele associated with endurance capacity with a greater proportion in all elite rugby athletes.	Chapter 2.3.2 Chapter 4
FTO rs9939609	0 = TT, 1 = AT & 2 = AA	A allele associated with greater mass and slightly greater in all elite rugby athletes.	Chapter 2.3.3 Chapter 5
COL5A1 *rs12722 and #rs3196378	*0 = TT, 1 = CT & 2 = CC #0 = AA, 1 = CA & 2 = CC	C allele of both SNPs associated with protection form soft tissue injury and in all elite rugby athletes.	Chapter 2.3.5 Chapter 7
APOE ε4 rs429358 and rs7412	0 = ε4+ & 2 = ε4-	ε4- associated with protection from poorer outcome following mTBI and is identified in 70% of elite rugby athletes.	Chapter 2.3.4 Chapter 6

Both *APOE* SNPs combine to generate the ε4 genotype (See chapter 2.3.4).

Table 13 Genotype score details for optimal **RU forwards** (GS_F) based on evidence from thesis observations (genotype score chosen from the data contained within the experimental chapters, 4-7).

Gene variant	Genotype score	Rationale	Evidence
ACE (I/D)	0 = DD, 1 = ID & 2 = II	I allele associated with athletic performance, particularly endurance performance and slightly greater proportion in all elite rugby athletes.	Chapter 2.3.1 Chapter 4
ACTN3 rs1815739	0 = RR, 1 = RX & 2 = XX	R allele associated with sprint/power performance. X allele associated with endurance capacity and with playing position elite rugby forwards.	Chapter 2.3.2 Chapter 4
FTO rs9939609	0 = TT, 1 = AT & 2 = AA	A allele associated with greater mass and elite rugby forwards.	Chapter 2.3.3 Chapter 5
COL5A1 *rs12722 and # rs3196378	*0 = TT, 1 = CT & 2 = CC #0 = AA, 1 = CA & 2 = CC	C allele of both SNPs associated with protection form soft tissue injury and in all elite rugby athletes.	Chapter 2.3.5 Chapter 7
APOE ε4 rs429358 and rs7412	0 = ε4+ & 2 = ε4-	ε4- associated with protection from poorer outcome following mTBI and is identified in 70% of elite rugby athletes.	Chapter 2.3.4 Chapter 6

Both *APOE* SNPs combine to generate the ε4 genotype (See chapter 2.3.4).

Table 14 Genotype score details for optimal **RU backs** (GS_B) based on evidence from thesis observations (genotype score chosen from the data contained within the experimental chapters, 4-7).

Gene variant	Genotype score	Rationale	Evidence
ACE (I/D)	0 = DD, 1 = ID & 2 = II	I allele associated with athletic performance, particularly endurance performance and slightly greater proportion in all elite rugby athletes.	Chapter 2.3.1 Chapter 4
ACTN3 rs1815739	0 = XX, 1 = RX & 2 = RR	R allele associated with sprint/power performance and with playing position elite rugby backs. X allele associated with endurance capacity.	Chapter 2.3.2 Chapter 4
FTO rs9939609	0 = AA, 1 = AT & 2 = TT	T allele associated with elite rugby backs.	Chapter 2.3.3 Chapter 5
COL5A1 *rs12722 and # rs3196378	*0 = TT, 1 = CT & 2 = CC #0 = AA, 1 = CA & 2 = CC	C allele of both SNPs associated with protection form soft tissue injury and in all elite rugby athletes.	Chapter 2.3.5 Chapter 7
APOE ε4 rs429358 and rs7412	0 = ε4+ & 2 = ε4-	ε4- associated with protection from poorer outcome following mTBI and is identified in 70% of elite rugby athletes.	Chapter 2.3.4 Chapter 6

Both *APOE* SNPs combine to generate the ε4 genotype (See chapter 2.3.4).

Data analysis

SPSS for Windows version 22 (SPSS Inc., Chicago, IL) software was used to conduct Pearson's Chi-square (χ^2) tests to compare genotype frequencies between athletes and controls, and between RU subgroups based on playing position (data presented in chapters 4-7). Differences in TGS between athletes and controls were analysed using unpaired t tests. Additionally, area under the ROC curves (AUC) were used to estimate the sensitivity of TGS to detect differences between backs and forwards (Zweig & Campbell, 1993). Alpha was set at 0.05.

8.1.3 Results

There was 100% agreement among reference samples genotyped in the three genotyping centres, i.e. Glasgow, Northampton and MMU laboratories. All genotype frequencies were in Hardy-Weinberg equilibrium for both the athlete and control group (Chapters *4*, *5*, *6* and *7*). Athletes were taller and heavier ($P < 0.05$) but not older ($P > 0.05$) than controls. No athletes had a TGS of zero or 100, regardless of GS model (range 8-92), two controls had a TGS of zero and one had a TGS 100 for the forwards GS model (Table 13).

Following the GS_{RU} model, the control group (TGS = 47.4 ± 16.8) had a lower mean TGS than all RU athletes (TGS = 49.9 ± 15.9; Figure 23**A**). Considering playing position and the GS_B model, TGS was greater (TGS = 57.1 ± 14.8) compared to controls (TGS = 53.4 ± 16.0; Figure 23**B**). Similarly, for the GS_F model the TGS was greater than that of the control group (TGS = 50.5 ± 14.8; TGS = 47.4 ± 16.8; Figure 23**C**).

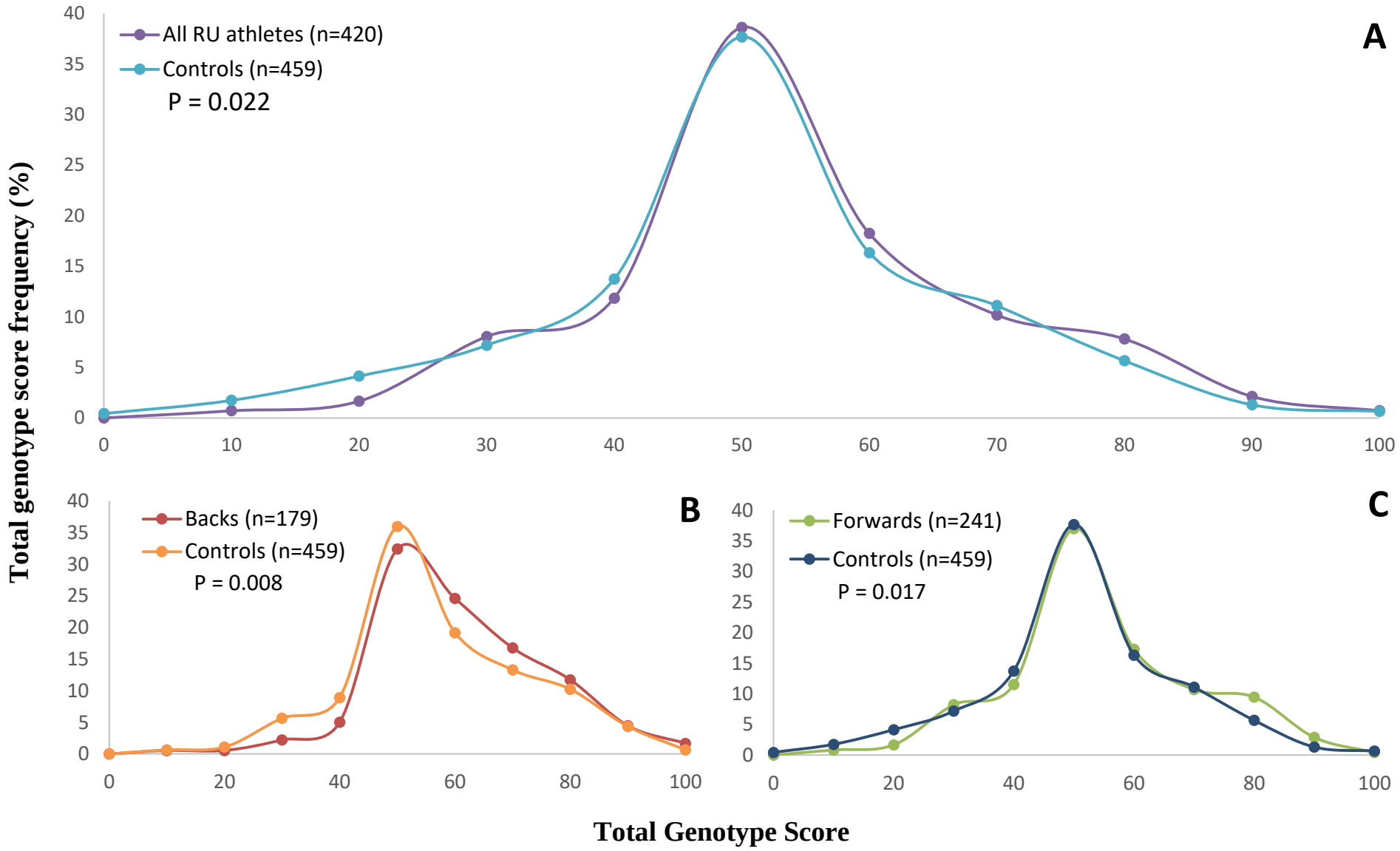


Figure 23 Frequency distribution of total genotype score. **A** All RU athletes (kurtosis = -0.309, SE = 0.238) versus controls (kurtosis = -0.098, SE = 0.227) as calculated from table 12. **B** Backs (kurtosis = 0.254, SE = 0.365) versus controls (kurtosis = -0.298, SE = -0.227) as calculated from table 14. **C** Forwards (kurtosis = -0.474, SE = -0.312) versus controls (kurtosis = -0.908, SE = -0.227) as calculated from table 13.

Of the backs group, 60% had a TGS greater than or equal to 60, whereas only 40% of the forwards had a TGS greater than or equal to 60, for their respective GS models (Figure 23). Furthermore, utilising the TGS_B/TGS_F ratio, the backs showed a greater ratio than forwards (1.27 versus 1.12), with ROC analysis identifing significant, but low, discrimination accuracy for TGS (AUC = 0.587, 95% CI 0.532-0.642, P = 0.002; Figure 24).

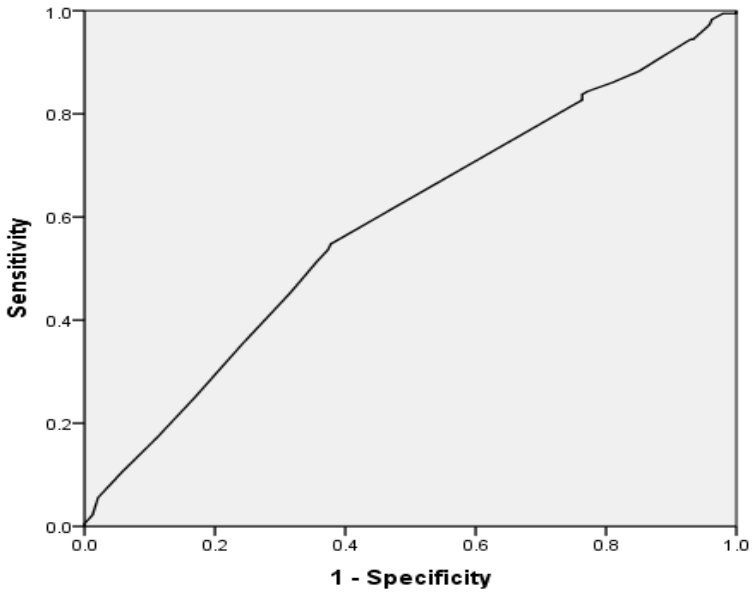


Figure 24 ROC curve for the TGS_B/TGS_F ratio (mean sensitivity = 0.571, mean specificity = 513).

8.1.4 Discussion

The present observations are the first to identify polygenic differences between elite RU athletes and a control group (Figure 23). By using the observed genotypic data to generate position specific GS, the present results show that the TGS of all athletes, backs and forwards were higher than their respective GS specific control groups (Figure 23). Furthermore, considerable polygenic differences were identified between backs and forwards, with the backs exhibiting a greater TGS and TGS_B/TGS_F ratio than forwards (Figure 24), for the six investigated genotypes.

The most unique aspect of the present chapter is the identification that the SNPs selected for inclusion in the present TGS models show greater predictability for the backs playing position (considering the results from chapter *4* and *5*; Figure 24). Addition of many other variants are necessary to identify a more forwards oriented or positionally ubiquitous TGS models. Nonetheless, the present observations have identified significant polygenic sensitivity in discriminating between elite RU backs and forwards, with the backs exhibiting the greatest TGS_B/TGS_F ratio (AUC = 0.587; Figure 24). These results highlight the combined genetic impact of the observed performance (Chapter *4*), anthropometric (Chapter *5*), neurobiological (Chapter *7*) and soft tissue (Chapter *6*) injury susceptibility observations, in relation to elite RU playing position.

The present data are in agreement with other TGS investigations of elite athlete status, in that the TGS of athlete are higher than that of controls, regardless of the GS model (Ruiz *et al.*, 2010; Eynon *et al.*, 2011b; Drozdovska *et al.*, 2013; Egorova *et al.*, 2014; Ahmetov *et al.*, 2015; Ben-Zaken *et al.*, 2015; Grealy *et al.*, 2015). However, some have not identified an association with TGS and elite athlete status (Gómez-Gallego *et al.*, 2010; Miyamoto-Mikami *et al.*, 2016). In relation to team sport specific analysis, Egorova et al. investigated polygenic profiles in a mixed athletic level soccer cohort (21% elite) and identified positional differences when compared to controls. However, they did not consider differences between playing positions (Egorova *et al.*, 2014), using the ratio ROC model presented in the current chapter and in Ben-Zakin et al. (2015) would have been an interesting addition to the study. Similarly, the present results have demonstrated a positional specific polygenic profile for both the forwards (TGS = 50.5) and to a greater extent in the backs (TGS = 57.1), compared to their respective controls (Figure 23**B** and 23**C**) and support the use of TGS in team sport genetic cohort studies.

TGS is the best currently available tool to understand the collective importance of multiple genetic variants on a given phenotype within a population. The present chapter is the first to utilise this statistical model in the context of athlete positional variation and its use was successful in discriminating between elite RU backs and forwards. It does however have a number of limitations which are inherent in its application. For example, the present chapter included a small number of genes, of which in isolation only two (*ACTN3* and *FTO*) were found to have an effect on playing position. The present models included all six studied gene variants within the TGS despite the association with playing position, as this was the original intention of the Williams and Folland method (Williams & Folland, 2008). However, the present chapter followed the data-driven approach to TGS and in a unique context of positional variation, where both homozygotes for both the *ACTN3* and FTO were found to be advantageous in different positions (R and T for backs and the X and A for forwards, respectively; Table 13 and 14). As such, different models were generated and were successful in identifying these individual positional group relationships - importantly in the presence of all six SNPs.

The inevitable addition of further SNPs to the elite rugby TGS model will improve the accuracy and confidence in the identified results and further development of these models is important for future investigations. Moreover, the present data show the importance of assessing positional specificity in terms of the allocation of the optimal genotype (Tables 12, 13 and 14). As identified in chapter *4* and *5*, considerable genetic variations exist within elite RU playing position that require multiple GS models to understand the genetic complexity of elite rugby athletes. The present chapter has identified TGS variation between backs and forwards (Figure 24), however RU consists of five distinct positional groups (Chapter *3.1.5*) and evaluation of TGS in each of these groups is the next step towards understanding the molecular aspects of elite rugby positional variation. This is one of the long-term goals of the RugbyGene project and an exciting prospect as the elite rugby sample continues to grow.

Chapter 9

Thesis general discussion

A portion of part *9.1.7* of this chapter is published in:

Heffernan S. M., Kilduff L. P., Day S. H., Pitsiladis Y. P. & Williams A. G. (2015): Genomics in rugby union: A review and future prospects. *European Journal of Sport Science*, 15(6), 460-468.

9.1 General Discussion

9.1.1 Overview

The capacity to achieve elite athletic success is known to be highly heritable (De Moor *et al.*, 2007), however attempting to explain this heritability has proven to be extremely difficult. This is partly because of the vast number of human molecular genetic variants currently known to exist (Sachidanandam *et al.*, 2001; International HapMap Consortium, 2005; Frazer *et al.*, 2007), but also due to the difficulty in identifying the heritable component of individual physiological traits that contribute to athletic success (Chapter *2.2.1*; Table 3). Currently, more than 155 genetic markers have been identified in association with athletic performance (Ahmetov *et al.*, 2016) and considerable effort has been devoted to understanding these molecular traits in recent years - often involving sprint/power or endurance athletes (Rankinen *et al.*, 2001; Berman & North, 2010; Eynon *et al.*, 2011c; Hughes *et al.*, 2011; Puthucheary *et al.*, 2011; Wilber & Pitsiladis, 2012; Eynon *et al.*, 2013a; Ma *et al.*, 2013; Pitsiladis *et al.*, 2013; Tucker *et al.*, 2013; Bouchard, 2015; Loos *et al.*, 2015) – chapter *2.2.4*. To date, only a small proportion of these efforts have considered the genomic component of team sport athletes, such as elite RU athletes, and often with considerable methodological limitations (Chapter *2.2.4*). The current thesis attempted to begin bridging the gap in the current understanding of team sport genomic variation, focusing on elite RU athletes. As such, the aims of the present research project were;

- To recruit a large biobank of elite rugby athletes for the purpose of evaluating the molecular genetic component of elite rugby status and investigate the molecular underpinnings of the physiological and anthropometric variation that exists in elite rugby playing position (Chapter *3.1.1*).

- To investigate *ACE* I/D and *ACTN3* R577X genotype distribution in elite rugby athletes. It was hypothesized that the *ACTN3* R allele and the *ACE* I allele would be more frequent in rugby athletes than controls. It was further hypothesized that *ACTN3* XX and *ACE* II genotypes would be underrepresented in backs compared to forwards, due to differences in overall work-to-rest ratio and differing requirements for high maximum speed (Chapter *4.1*).

- To investigate if FTO rs9939609 genotype differs between elite rugby athletes and a control population, and/or between playing positions. Based on prior data in obese populations, it was firstly hypothesised that the rs9939609 risk (A) allele would be would be overrepresented in playing positions typically requiring greater body and muscle mass, while the protective (T) allele would be more common in positions requiring a lean phenotype (Chapter *5.1*).

- To quantify the 'at risk' *APOE* ε4 carriers in the elite rugby community and to investigate if *APOE* genotypes differed between elite rugby athletes and a control population. Based on the published *APOE* ε4/ε4 association with poorer outcome following brain injury, it was hypothesised that the ε4/ε4 genotype and ε4 allele would be underrepresented in elite rugby athletes compared to controls (Chapter *6.1*).

- To investigate if associations of *COL5A1* rs12722 and rs3196378 genotype and allele frequencies differed between elite rugby athletes and a control population. It was hypothesised that the *COL5A1* rs12722 and rs3196378 protective C alleles and CC genotypes would be overrepresented in elite rugby athletes compared to controls (Chapter *7.1*).

- Finally, a TGS algorithm was applied to assess the polygenic profile, of the gene variants examined in the first four experimental chapters of the present thesis, for all RU athletes, forwards and backs, compared to controls (Chapter *8.1*).

9.1.2 Main experimental findings

The main findings of the current thesis are discussed, in detail, in the subsequent sections of this chapter. However, briefly, the R allele of *ACTN3* R577X rs1815739 was identified as advantageous for backs, particularly the back three players, reflecting their positional requirements (Chapter *4.1*). However, the results do not support *ACE I/D* as a genetic marker for rugby performance, showing no differences between athletes and controls or positional groups. These chapters demonstrated the value of single sport cohorts, particularly in team sports, and the need for large sample sizes when conducting candidate gene association studies in sport. Similarly, the T allele of *FTO* rs9939609 may be beneficial to elite athletes who rely greatly on lean mass relative to total body mass for athletic success (RU back three and centre players; Table 1) as they are more likely to carry an obesity protective T allele and may ultimately be selected for the appropriate playing positions as a result (Chapter *5.1*). The *APOE* ε4 genotype appears to shows no genotypic or allelic advantage for elite rugby athlete status. However, the data do show a considerable number of elite rugby athletes exist who possess one or two *APOE* ε4 alleles and these athletes may be at greater risk of poorer outcome following an mTBI event (Chapter *6.1*). This particular result requires considerable future investigation as the possible implications of ε4 carriage may result in severely reduced neurological function following an athletes playing career. Chapter *7.1* presented the first association between the *COL5A1* 3′ UTR rs12722 and rs3196378 and elite competitive status in a large cohort of athletes. The C alleles of both polymorphisms were overrepresented in all RU athletes, forwards and backs versus controls. Finally, incorporating all six of the aforementioned polymorphisms into a polygenic profile revealed differences between elite RU athletes and controls. By using ROC analysis of TGS_B/TGS_F ratio the present results identified significant, but low, discrimination accuracy for TGS as a predictor of RU playing position (Chapter *8.1*).

As previously mentioned in chapter *1*, RU athletes are different from other athlete groups, in that there are vast differences in the physiological and anthropometric characteristics across a single RU team according to playing position (Smart *et al.*, 2013; Table 1). RU is also distinctive as individual clusters of positions require different movement patterns (Quarrie *et al.*, 2013) and thus differ in their metabolic demands. In terms of positional specific physiological differences that may be reflected in players' genetic variation, backs show lower maximal strength compared to forwards in terms of bench press, back squat and power clean (Smart et al., 2014). However, backs are faster, sprinting 10 m and 20 m than forwards (Smart *et al.*, 2014) and these differences become larger when specific positions are considered (Smart *et al.*, 2013). Furthermore, meta-analyses have shown that for every 1000 hours, an elite RU athlete will experience approximately 81 injuries during match play and three during training, with the majority being ligament, tendon and muscle injuries of the lower limbs (Williams *et al.*, 2013). Indeed, in the most recent Rugby World Cup (2015) this rate of incidence was more than 90 injuries per 1000 h (Fuller *et al.*, 2016). Furthermore, injury incidence differs across RU playing position, with elite back row players showing the highest rate among forwards and centres the highest among backs (Fuller *et al.*, 2016). Current mTBI incidence for elite RU players ranges between 4.6-8.9 per 1000 playing hours and has grown over time (Gardner *et al.*, 2014; Cross *et al.*, 2015; Fuller *et al.*, 2015a). In relation to head injury, the most recent Rugby World Cup data (2015) shows that indicates of mTBI were the most common injury (14%; n = 24) and accounted for a total of 184 days absence (mean ~8 days each) from training and competition during and after that tournament (Fuller *et al.*, 2016). These data are reflected, to some extent, in the genetic data observed in the present thesis and the following three sections will discuss these genetic findings, that may help in explaining some of these physiological quantities (Chapter *2.2.1*) and existing injury rates (Chapter *2.2.3*) that may predisposes rugby athletes to elite success. Following achievement of the first thesis aim (Chapter *2.3.6*), the subsequent aims (Chapters) had three consistent themes; (1) comparing the genetic profiles of 'all' elite RU athletes to controls,

(2) identifying genetic difference between elite RU playing position and controls and (3) comparing the genetic profiles of elite RU backs to elite RU forwards.

9.1.3 Genetic profile of elite RU athletes compared to controls

Given the considerable anthropometric and physiological differences between elite RU athletes and the general population (Table 1), it was hypothesised that their respective genetic profiles would differ. The results of the present thesis shows that this is heavily dependent on the particular genetic variant being investigated.

The results presented in chapter 7 show an association between *COL5A1* 3′ UTR rs12722 and rs3196378 and elite RU athlete status. Specifically, the C alleles of both polymorphisms were overrepresented in all RU athletes versus controls. These results provide an insight into the potential injury susceptibility of some elite rugby athletes. As such, both rs12722 and rs3196378 have previously been associated with tendinopathy in Australian Caucasians, but only the former in South African Caucasians (September *et al.*, 2009). Accordingly, rs12722 has also shown association with flexibility (Collins *et al.*, 2009; Lim *et al.*, 2015), exercise-associated muscle cramping (O'Connell *et al.*, 2013), anterior cruciate ligament injury (Posthumus *et al.*, 2009b; Altinisik *et al.*, 2015; O'Connell *et al.*, 2015) and Achilles tendinopathy (Mokone *et al.*, 2006). In these studies, the minor CC genotype was overrepresented in the respective asymptomatic controls and suggests a protective role of the C allele against injury.

Considering the high frequencies of tendon and ligament injuries in elite RU (Williams *et al.*, 2013; Fuller *et al.*, 2015a; Fuller *et al.*, 2016), the present results that the injury protective C alleles were more representative of elite RU athletes than controls (Figure 22 and Table 10), suggests a trend towards selection of athletes that are less likely to accrue injury, in

terms of their less frequent availability to coaches. Indeed, RU has one of the highest reported rate of injury incidence in professional team sports (Brooks & Kemp, 2008), with the most recent Rugby World Cup (2015) data indicating an incidence rate of more than 90 injuries per 1000 playing hours (Fuller *et al.*, 2016). Therefore, RU athletes possessing more C alleles at these two genetic loci are probably at a lower risk of injury, given their exposure to the high-risk environment of elite rugby.

Continuing with possible injury susceptibility, no difference was identified in *APOE* ε4 frequency between athletes and controls. However, importantly, ~30% of the present elite rugby athlete cohort were identified as risk allele carriers (ε4+). As mentioned in chapter 6, ε4+ rugby athletes may be at greater risk of cognitive and physical impairments following mTBI, compared to non-carriers (Lichtman *et al.*, 2000; Crawford *et al.*, 2002; Zhou *et al.*, 2008; Müller *et al.*, 2009; Noé *et al.*, 2010; Roozenbeek *et al.*, 2013; Zeng *et al.*, 2014; Lawrence *et al.*, 2015; Li *et al.*, 2015; Merritt & Arnett, 2016). This could have widespread practical implications because World Rugby (the international governing body of rugby union) estimates there are 7.23 million rugby players worldwide (*http://www.worldrugby.org/development/player-numbers*). Thus, over two million ε4+ rugby players may be at greater risk of poorer outcome following mTBI than their ε4- counterparts. As Cross et al. (2015) has recently shown a 60% greater chance of time-loss injury in elite players who returned to competition in the same season as receiving a mTBI, correctly managing athletes that are at greater risk due to any predisposing factor such as the 30% who are ε4+, could become a valuable strategy for researchers, support scientists and medical staff in due course. In addition to these possible short-term effects of repeated injury, carriage of the ε4 allele could have considerable long-term neurological effects on retired players, later in life, and more scientific attention needs to be focused on these athletes.

The *ACE* I/D polymorphism is one of the most studied variants in the context of human performance, Ma *et al.* (2013) reported that the *ACE* II genotype was associated with physical performance (OR = 1.23), with the *ACE* I allele being associated with elite endurance performance in a variety of events (Gayagay *et al.*, 1998; Montgomery *et al.*, 1998; Myerson *et al.*, 1999; Alvarez *et al.*, 2000; Cieszczyk *et al.*, 2009; Cieszczyk *et al.*, 2010) and the D allele associated with superior performance in sprint and power-related sports (Woods *et al.*, 2001; Kikuchi *et al.*, 2012; Eider *et al.*, 2013; Papadimitriou *et al.*, 2016). However, the current thesis reports no *ACE* I/D genotype difference between elite rugby athletes and controls (Chapter *4*). This lack of association contrasts with the afore mentioned meta-analysis where the *ACE* I allele was associated with physical performance (Ma *et al.*, 2013). Nevertheless, given the mixed metabolic nature of RU (Chapter *2.1.2*), a comparable association in the present thesis was less likely. The importance of *ACE* I/D remains controversial in the literature, with no associations reported in other isolated team sports such as elite European soccer (Gineviciene *et al.*, 2014) and non-elite rugby athletes (Bell *et al.*, 2010). These prior data, in conjunction with the current findings in a larger study that also considers playing position, suggest that *ACE* I/D plays little role in performance of team sport athletes. *ACE* I/D genotype-athlete phenotype associations are more likely to exist in specialized endurance athletes (Puthucheary *et al.*, 2011).

Similar results were identified for *ACTN3 R577X* and *FTO*, in that no differences were observed for genotype or allele distribution between RU athletes and controls (Chapters *4*, *5* and *6*). For *ACTN3*, this result was particularly interesting as the RR genotype has previously been associated with speed and power performance (key components in elite RU success) in two independent meta-analyses (Alfred *et al.*, 2011; Ma *et al.*, 2013). However, this result was in agreement with a study considering many sports simultaneously (reviewed in detail in chapter *2.2.5.7*), where team sport athlete status showed no association with *ACTN3* R577X genotype (Eynon *et al.*, 2014) and in isolated team sports (Djarova *et al.*,

2011b; Ruiz *et al.*, 2011b; Bell *et al.*, 2012c; Egorova *et al.*, 2014; Garatachea *et al.*, 2014). Only one study has investigated *FTO* genotype in an athletic populations (detailed review in chapter *2.3.3*), in three European cohorts of power (n = 258; 58.3% elite) and endurance athletes (n = 266; 57.1% elite) from a variety of sporting disciplines, with no associations between *FTO* and athlete status identified (Eynon *et al.*, 2013b). The results of the present thesis are again in agreement (Chapter *5*) and again are surprising given that *FTO* genotype is associated with greater body mass (detailed review in chapter *2.2.3*) and elite RU athletes are on average ~25kg heavier than the general population (Chapter *2.1.1*, Table 1).

Although, individually, *ACE*, *ACTN3*, *FTO* and *APOE* ε4 did not differ between all RU athletes and controls, when the data from these variants was combined with *COL5A1* in a TGS (Chapter *8*), athletes showed a greater score than controls (Figure 23). This result was in agreement with other TGS investigations of elite athletes, despite any subjectivity in SNP choice and optimal genotype allocation (Ruiz *et al.*, 2010; Eynon *et al.*, 2011b; Drozdovska *et al.*, 2013; Egorova *et al.*, 2014; Ahmetov *et al.*, 2015; Ben-Zaken *et al.*, 2015).

9.1.4 Genetic profile of elite RU playing position compared to controls

While the lack of association between all RU athletes and controls may have been unforeseen in relation to *ACE*, *ACTN3*, *FTO* and *APOE*, it was not surprising considering positional variation that exists across all physiological (Chapter *2.1.1*, Table 1) and game demand variables (Chapter *2.1.2*, Table 1). It is possible that these considerable physical differences across RU playing positions are reflected in genetic characterises but those differences remain hidden when comparing the entire cohort to controls. Therefore, it was important to investigate the genetic differences between the positional subgroups (Chapter *3.1.1*) and the general population.

One of the most remarkable findings of the present thesis was the low frequency of the *ACTN3* XX genotype among the back three athletes (8.7%), approaching although not as low as the frequency observed in elite sprinters (Yang *et al.*, 2003; Niemi & Majamaa, 2005). As eluded to in chapter *4*, the XX genotype is present in ~18% of Caucasians (Table 6) and indicates absence of the α-actinin-3 protein (Beggs *et al.*, 1992; Mills *et al.*, 2001). Absence of α-actinin-3 could hinder back three (wing and full back) sprint ability. Interestingly, the X allele was overrepresented in forwards (52.5%) compared controls (42%), which is understandable because mouse models have shown a higher propensity for aerobic enzyme activity and greater force recovery after fatigue in α-actinin-3 deficient mice (Seto *et al.*, 2011; Seto *et al.*, 2013). This could indicate that XX genotype humans might have a greater capacity for recovery from fatiguing exercise - a trait which would benefit forwards with their more sustained match play intensity and necessity for quick recovery (Chapter *2.1.2*, Table 1). For *FTO* variation, there were fewer AA homozygotes and more T allele carriers (Chapter *5*, Figure 20A and Table 7) in back three and centres group than controls. Additionally, controls had more than twice the odds of being AA than the back three and centres group, with greater odds of T allele carriers in the back three and centres than controls (Table 7). The results of chapter *5* suggests an RU position specific advantage, in that the greatest T allele and TT genotype frequencies were in the RU athletes more reliant in a lean phenotype for success (Table 1; Smart *et al.*, 2013). Furthermore, the ability to rapidly produce high levels of power relative to body mass using the leg musculature is greater in those playing positions more reliant on a lean phenotype (Crewther *et al.*, 2012).

No distinct associations between playing position and controls were identified for *ACE*, *APOE* or *COL5A1* genetic variation, however when their combined influence (including *ACTN3* and *FTO*) were considered in chapter *8*, differences were evident. As detailed in chapter *8.1.2*, different GS models were generated to account for the divergence in optimal

polygenic profile between backs and forwards, as identified in chapters *4* and *5*. For both backs (TGS = 57.1 ± 14.8) and forwards (TGS = 50.5 ± 14.8), TGS was greater compared to their respective controls (TGS = 53.4 ± 16.0, TGS = 47.4 ± 16.8; Figure 23**B** and 23**C**). This result was in agreement with other TGS investigations of elite athlete status (Ruiz *et al.*, 2010; Eynon *et al.*, 2011b; Drozdovska *et al.*, 2013; Egorova *et al.*, 2014; Ahmetov *et al.*, 2015; Ben-Zaken *et al.*, 2015). In relation to team sports specifically, Egorova et al. investigated the polygenic profiles in a mixed athletic level soccer cohort (21% elite) and identified differences between controls and individual positional compared (Egorova *et al.*, 2014), with which the present findings are in agreement (Chapter *8.1.3*, Figure 23).

9.1.5 Variation in genetic profile of elite RU playing position

Continuing from the discussions in the previous sections (*9.1.2.1* and *9.1.2.2*), where genetic variation was identified between all RU athletes, playing position subgroups and controls, the next logical step is to discuss the inter-positional variation. As such, two SNPs showed differing genetic variation between playing positions (*ACTN3*, chapter *4* and *FTO*, chapter *5*).

The *ACTN3* XX genotype was almost twice (OR = 1.77) as common in forwards than backs, which suggests α-actinin-3 deficient individuals are more suited to forward play. Furthermore, the X allele was overrepresented in forwards (52.5%) compared to backs (37.8%; OR = 1.49), with the R allele frequency greater in the back three (68.8%) compared to forwards (47.5%; OR = 2.00) and the other backs (58.2%; OR = 1.59; Chapter *4*, Figure 19**B**). As mentioned above, the 69 back three athletes (wings and fullbacks) included only six individuals (8.7%) of XX genotype which differed from the forwards (24.8%) who were over three times (OR = 3.46) more likely to be XX genotype than the back three athletes. Furthermore, the remaining backs (centres and halves) were over twice as likely to show the

α-actinin-3 deficient genotype than the back three athletes (OR = 2.59). It appears as though the shorter match play rest periods for forwards compared with backs - work to rest ratios 1:7 and 1:22, respectively (Table 1; Deutsch *et al.*, 2007) indicates that greater fatigue resistance and would be particularly beneficial for forwards. Moreover, the greater calcineurin activity in XX homozygote humans and approximately threefold increase in calcineurin activity and distance run after endurance training in KO mice (Seto *et al.*, 2013), further support the notion that forwards would benefit from a greater fatigue resistance.

Regarding the R allele advantage in the back three players, R allele carriers have a greater proportion of type II and IIx fibres and larger relative surface area per IIx fibre than XX carriers (Vincent *et al.*, 2007; Ahmetov *et al.*, 2011; Broos *et al.*, 2012; Broos *et al.*, 2016). In fact, Broos et al. (2016) showed corresponding single fibre characteristics. Fibres of RR humans showed greater contractile velocity than XX individuals while exhibiting similar isometric force production. This suggests a strong rationale for the R allele advantage in elite sprinters (Yang et al., 2003; Niemi & Majamaa, 2005) and the back three athletes of the present chapter (Table 6 and Figure 19). This also supports the *ACTN3* R allele has a more relevant association with relative muscle power than absolute power (Kikuchi *et al.*, 2014b), due to preservation of force at high contractile velocities and not absolute force (Broos *et al.*, 2016), a crucial relationship for sprinting performance (Miller *et al.*, 2012; Morin *et al.*, 2012). As such, rugby athletes where relative muscle power is an important quantity (Crewther *et al.*, 2012) would have a greater advantage with position of the R allele, as shown in the present results (Table 6 and Figure 19). This would be of obvious benefit to the back three players as a fundamental requirement to selection in this position is sprint performance. Seto *et al* (2013) showed that the likely mechanism for this genotype-phenotype association is via the calcineurin muscle fibre remodelling pathway (for details see chapter *4.1.4*). This could explain the advantage of R allele carriers over α-actinin-3

deficient XX individuals for high velocity contractions – particularly important for back three players. These data suggest that the *ACTN3* R577X polymorphism has the potential to contribute to position-specific player profiling when combined with other genetic and physiological data in the future.

For *FTO* positional variation, the backs had a greater frequency of T allele carriers than forwards (Chapter *5.1.3*, Table 7, Figure 20) and showed greater odds of being T allele carriers than AA genotype (OR = 1.84, Table 8). Compared to forwards, TT genotype and T allele carriers were more common in the back three and centres group (Chapter *5.1.3*, Figure 20A and Table 7). Likewise, forwards had greater than three times the odds of being AA genotype than the back three and centres group with greater odds of T allele carriers in the back three and centres group than forwards (Table 8). This suggests an advantage for forwards that carry the A allele above those that do not, possibly a result of lower body mass associated with the T allele. As detailed in chapter *5*, until recently little was known about the molecular basis for *FTO* SNP associations with any reported phenotype measure because there was no association between *FTO* SNPs and expression of the FTO protein (Wåhlén *et al.*, 2008; Grunnet *et al.*, 2009). However, *FTO* has recently been found to influence IRX3 protein expression and individuals possessing the protective *FTO* genotype/allele (TT/T) display lower IRX3 expression than AA homozygotes (Smemo *et al.*, 2014), which is hypothesised to lead to long term motor unit availability (Chapter *5*). Furthermore, recent associations between *FTO* variants and IGF-1, specifically that serum IGF-1 levels were greater in T allele carriers (Rosskopf *et al.*, 2011), may provide an additional mechanism to explain the greater T allele frequency in the back three and centres who rely on lean muscle tissue for success (detailed in chapter *5*, Figure20). Similar to *ACTN3*, *FTO* variation has the potential to contribute to position-specific player profiling when combined with other

genetic and physiological data in the future and possibly lean muscle development (not yet experimentally tested).

Again, no distinct associations between RU playing position were identified for *ACE*, *APOE* or *COL5A1* genetic variation, however when their combined influence (including *ACTN3* and *FTO*) was considered in chapter *8*, differences were evident. Nonetheless, the present observations have identified significant polygenic sensitivity in discriminating between elite rugby union backs and forwards, with the backs exhibiting the greatest TGS_B/TGS_F ratio (AUC = 0.587, Figure 24). This is an important advancement in identifying the molecular characteristics of RU playing position and as more SNPs are included the sensitivity of the discrimination will increase. These developments may lead to the accurate use of genetic information for positional identification, individualised training and injury management in the future.

9.1.6 Study limitations

The present research project began with a review of the methodological limitations present in studies of a similar design (Chapter *2.2.4*) and considerable efforts were made to limit any identified concerns. However, some limitations exist within the present thesis despite the greatest effort to 'control the controllables' and are discussed below.

Although genetic case-control association studies, such as the present thesis, have achieved great successes in identifying some of the genetic component of complex exercise and sport related phenotypes (Rankinen *et al.*, 2001; Loos *et al.*, 2015; Ahmetov *et al.*, 2016), case-control association studies only investigate variation at a select and often small number of loci, which only explains a limited proportion of heritability of athlete status (Pitsiladis *et al.*, 2013; Tanaka *et al.*, 2016). The present thesis only investigated seven SNPs, generating six genotypes, however each one of these genotypes was chosen based on their relevance to

the study sample and all rationale for inclusion in the present thesis were well evidenced in the scientific literature (Chapter *2.3*). Nevertheless, including a small number of SNPs of the ~155 identified in the context of athletic performance (Ahmetov *et al.*, 2016) and in the shadow of the many yet to be identified variants (Chapter *2.2.2*) remains a limitation to the present thesis. The fundamental reason for the inclusion of only seven SNPs was time and financial resources, caused by the difficulty in recruiting and gaining access to cohorts of elite professional RU players – a process that was often extremely time consuming and incurred considerable financial commitments. However, while these time and budgetary costs meant the inclusion of few SNPs, the present sample (RugbyGene Project) to the author's knowledge, is currently the largest (100% elite) single sport athletic cohort in sports genomics to date, as part of the GENESIS cohort (for a review of other athlete cohorts see Pitsiladis *et al.*, 2016).

Secondly, the nature of genetics research dictates that extremely large sample sizes are needed before genuine conclusions can be made about the variation within a given cohort, for a given phenotype (Gauderman, 2002). However, due to the natural rarity of elite athletes it is difficult to congregate large samples of particular athlete groups and even more difficult to recruit large cohorts of single sport athletes (such as RU players). Consequently, large numbers (many hundreds) of 'Tier 1' rugby athletes are required – and was achieved through international research collaboration. While the present sample of ~450 elite Caucasian RU athletes is large by sports genomics standards (Chapter *2.2.4*), confidence in the present results will be enhanced with a greater sample size and conducting larger analyses such as genome wide association studies (GWAS) requires additional samples. Regardless, for the purpose of the present thesis, the current sample was amply powered (Hong & Park, 2012).

9.1.7 Directions for future research

9.1.7.1 Genotype-phenotype associations that should be investigated

Future objectives of the RugbyGene project (based on the present thesis) within the broader Athlome project (Pitsiladis *et al.*, 2016) includes investigating whether additional genetic variants associated with excellence in other sports are similarly associated in the multifaceted sport of rugby. The next step would be to investigate genotype associations with a range of phenotypes such as, physiological, anthropometric, other performance variables and incidence of injury, etc.

Identifying genetic associations (in a cohort of elite RU athletes) with rugby-specific physiological and anthropometric variables, for example those aspects of strength, speed and body composition assessed by Smart et al. (2013), would further exemplify the importance of the genetic component to RU. Furthermore, rugby union has one of the highest reported incidents of match play injuries in all professional team sports (Brooks & Kemp, 2008), with an injury incidence of ~90 per 1000 match play hours resulting in ~30 days absence per injury (Fuller *et al.*, 2016). Research collaborations that combine these kinds of large, meticulously-collected rugby injury databases with genetic analyses conducted on those very same players, could be extremely fruitful in explaining some of the as yet unexplained inter-individual variability in injury susceptibility and could identify novel markers of injury risk in rugby.

Concussion risk in rugby (Chapter *2.1.4*), as well as consideration of the potential longer-term consequences, is clearly and justifiably a topic of much attention at this time (Raftery, 2013; Fuller *et al.*, 2015b; Raftery *et al.*, 2016). Accordingly, Gardner et al. (2014) recently

conducted a comprehensive meta-analysis showing that RU, the incidence rate of concussion was ~4.7 per 1000 match play hours. Data from 16 studies showed that at the elite/international level the incidence rate was lower (~1.2 per 1000 match play hours), though still considerable (Gardner *et al.*, 2014). Furthermore, the most recent Rugby World Cup data (2015) identified mTBI as the most commonly occurring injury (14%) and accounted for 184 days absence from training and competition (Fuller *et al.*, 2016). These notable injury rates make research efforts to identify molecular markers for the risk of specific rugby-related injuries, such as concussion, highly warranted. Indeed, the development of tools to identify individuals at greater risk of concussion in rugby and greater risk of longer-term pathological neurobiological changes following a career playing rugby would seem highly responsible in the context of player welfare. Following the demonstration that ~30% of rugby athletes may have a higher risk of poorer outcome follow brain injury, with only one genetic variant (Chapter *6*), more risk variants need to be analysed for their presence in elite rugby athletes. Importantly, these data then need to be combined with concussive incidence, recovery and biochemical markers.

9.1.7.2 Advanced genomics technology

For RU as in other areas within sports genomics, the starting point for the investigation of genetic variation is through the candidate gene (hypothesis driven) approach, such as the present thesis. This approach of considering the biological mechanisms of a given trait (for example, incidence of tendinopathy or high VO_{2max}) and investigating previously identified genetic variants within genes known or suspected to affect the relevant biological pathways is valuable initial analysis. While the candidate gene approach is by far the most utilised technique in quantifying molecular genetic markers of sport-related phenotypes and a good starting point, other more complex analysis techniques are required for the future of this field to realise its full potential (Pitsiladis *et al.*, 2013).

The hypothesis-free approach of conducting GWAS has been utilised extensively to identify new genetic variants in various domains within human biology (Wolfarth *et al.*, 2014) and is recommended for identifying novel genetic variants in RU. GWAS is the process of investigating large numbers of known SNPs simultaneously (~2 million, for example) for a given complex trait (Visscher *et al.*, 2012). As already mentioned, complex traits of relevance to RU could be sprinting ability, muscle strength, incidence of injury or simply being an elite rugby union player (elite athlete status). Importantly, given the large number of hypotheses tested statistically, only the strongest associations are usually accepted to be true results (e.g. when $P < 5 \times 10^{-8}$), although a very large cohort size and/or strong genetic effect sizes are usually required for this approach to be effective. A strength of any hypothesis-free approach like GWAS is that new variants which reveal new biological insight can be discovered - and then further investigated experimentally.

Genetic testing technologies have advanced to such an extent that investigating all nucleotides in a gene, all protein-coding genes, or even the whole genome is now possible via direct sequencing. Depending on the availability of funding, participants and other resources including laboratory equipment and bioinformatics expertise, multiple options are available to exploit these techniques for the identification of novel mutations, polymorphisms or structural variants. Firstly, targeting specific genes associated with a given phenotype, such as the *MSTN* gene and muscle mass (Schuelke *et al.*, 2004), and sequencing every nucleotide (~7,000 bp) of that gene in a large cohort of elite RU players for whom strength and muscle mass are also known could be an elegant approach to identifying novel genetic variants associated with muscle size and strength in rugby players. To examine larger regions than one gene, one could utilise whole-exome sequencing, which uses the same rational as targeted gene sequencing but to a much greater extent, targeting all protein-coding DNA sequences (~230,000 exons or ~30 million bp; ~1% of the human

genome). Finally, providing that considerable resources including an appropriately large cohort were available, assessment of the whole-genome (~3 billion bp) would be the ultimate, most comprehensive method of identifying novel mutations or polymorphisms of functional importance in RU athletes. Currently, whole-genome-sequencing is not possible for reasons of cost, logistics, interpretation and statistics, although the eventual application to rugby is inevitable.

9.2 Conclusion

The results from the present thesis identify considerable genetic variation in a relatively large cohort of 100% elite RU athletes. With only a few SNPs, the thesis results show the existence of genetic variation across team sport playing position - which is often overlooked in the attempt to increase sample size. Furthermore, two of the genes studied in the present thesis related directly to the injury susceptibility of elite players (*COL5A1* and *APOE*). This knowledge is the first step towards utilising genetic information to improve player management and welfare - following further research. In fact, the observation that ~30% of players may be at greater risk of poorer neurological recovery following mTBI requires direct attention from rugby governing bodies and the related science and medical communities. Nevertheless, the present thesis was the first to identify these genetic associations and should be seen as initial work for future research to build upon in order to progress the field of genetic analysis in sport.

Appendices

Heffernan *et al. BMC Genetics* (2017) 18:4
DOI 10.1186/s12863-017-0470-1

BMC Genetics

RESEARCH ARTICLE Open Access

Fat mass and obesity associated (*FTO*) gene influences skeletal muscle phenotypes in non-resistance trained males and elite rugby playing position

S. M. Heffernan[1*], G. K. Stebbings[1], L. P. Kilduff[2], R. M. Erskine[3,4], S. H. Day[1], C. I. Morse[1], J. S. McPhee[1,2], C. J. Cook[2,6], B. Vance[7], W. J. Ribbans[5], S. M. Raleigh[5], C. Roberts[8,9], M. A. Bennett[2], G. Wang[10], M. Collins[11], Y. P. Pitsiladis[10] and A. G. Williams[1,4]

Abstract

Background: *FTO* gene variants have been associated with obesity phenotypes in sedentary and obese populations, but rarely with skeletal muscle and elite athlete phenotypes.

Methods: In 1089 participants, comprising 530 elite rugby athletes and 559 non-athletes, DNA was collected and genotyped for the *FTO* rs9939609 variant using real-time PCR. In a subgroup of non-resistance trained individuals (NT; $n = 120$), we also assessed structural and functional skeletal muscle phenotypes using dual energy x-ray absorptiometry, ultrasound and isokinetic dynamometry. In a subgroup of rugby athletes ($n = 77$), we assessed muscle power during a countermovement jump.

Results: In NT, TT genotype and T allele carriers had greater total body (4.8% and 4.1%) and total appendicular lean mass (LM; 3.0% and 2.1%) compared to AA genotype, with greater arm LM (0.8%) in T allele carriers and leg LM (2.1%) for TT, compared to AA genotype. Furthermore, the T allele was more common (94%) in selected elite rugby union athletes (back three and centre players) who are most reliant on LM rather than total body mass for success, compared to other rugby athletes (82%; $P = 0.01$, OR = 3.34) and controls (84%; $P = 0.03$, OR = 2.88). Accordingly, these athletes had greater peak power relative to body mass than other rugby athletes (14%; $P = 2 \times 10^{-6}$).

Conclusion: Collectively, these results suggest that the T allele is associated with increased LM and elite athletic success. This has implications for athletic populations, as well as conditions characterised by low LM such as sarcopenia and cachexia.

Keywords: RugbyGene project, IRX3, Lean mass

Background

Fat mass and obesity associated (*FTO*) is the most investigated gene in obesity and has complex molecular mechanisms that are yet to be elucidated. Recent genome-wide association studies (GWAS) have identified several common single nucleotide polymorphisms (SNP) in the human *FTO* gene associated with obesity, body mass index BMI; [1], cardiovascular disease and hypertension [2, 3]. These *FTO* SNPs, which are in strong linkage disequilibrium ($r^2 > 0.80$), are located in a cluster on the first intron of the gene on chromosome 16 and consequently exhibit similar obesity-related traits [4]. Thus, within different *FTO* variants, those alleles that have been positively associated with obesity-related phenotypes are referred to as risk alleles, while those negatively associated with such traits are referred to as protective alleles. Homozygotes for the minor risk allele consistently demonstrate greater

* Correspondence: shane.heffernan@stu.mmu.ac.uk
[1]MMU Sports Genomics Laboratory, Manchester Metropolitan University, Crewe Green Road, Crewe CW1 5DU, UK
Full list of author information is available at the end of the article

BMI and body mass (3–10 kg) in comparison to protective allele carriers [5, 6]. This greater body mass is likely to be adipose tissue [7–11], although there exist some suggestions of greater lean mass (LM) in addition to fat mass [9, 12] and independent of fat intake and physical activity [9]. This suggests that *FTO* genotype may be related to muscle properties and is supported by evidence from a large UK twin study that related *FTO* SNPs with body composition while controlling for lean mass and fat mass (separately and combined). The authors concluded that *FTO* SNP associations with body size were a composite of both lean and fat mass, not fat mass alone [13].

Environmental lifestyle factors (diet and physical activity) have also been investigated for *FTO* gene-environment interactions. Risk allele carriers are more likely to choose a high fat diet than protective allele carriers [11, 14, 15]. However, with administration of a high protein diet (25% energy intake) risk allele carriers demonstrated a greater reduction in body mass, fat mass and percentage body fat [16], due to greater appetite suppression than in protective allele carriers [17]. Additionally, physically active risk allele carriers have a 30% reduction in likelihood of becoming obese and have 36% less body fat compared to inactive risk allele carrying individuals [18]. In contrast, data from the HERITAGE Family Study showed that following 20 weeks of endurance training, protective allele carriers exhibited reductions in fat mass three times greater than risk allele homozygotes [19]. Interestingly, when comparing normal weight and obese individuals who participate in sport, no differences in *FTO* genotype were observed ($P = 0.97$), which was contrasted by those not participating $P = 0.02$; [20]. Considering the attenuation of *FTO*-associated obesity with environmental factors and the greater *FTO*-associated LM reported in obese populations [9, 12], investigating LM and associated phenotypes in healthy, non-obese, non-resistance trained individuals and habitually trained elite athletes would be worthwhile.

To date, there have been no investigations of in vivo skeletal muscle phenotypes in trained or non-resistance trained populations for associations with *FTO* genotype. Eynon et al. [21] investigated *FTO* rs9939609 in three European cohorts of power ($n = 258$; 58.3% elite) and endurance athletes ($n = 266$; 57.1% elite) from a variety of sporting disciplines - but identified no associations. This lack of association was likely due to the considerable differences in physiological demand between the various athletic disciplines included, plus further variability in the standard of athlete. We have recently shown the ability of genetic research in a single sport with player roles that differ distinctly, namely rugby union (RU), to reveal context-specific competitive advantages provided by particular alleles [22]. Therefore, as RU includes athletes of remarkably distinct anthropometric and body composition phenotypes, elite RU provides a unique opportunity to investigate *FTO* in individuals at the extreme upper end of physical fitness [23].

Therefore, the main aims of the present study were to (1) investigate any association of *FTO* rs9939609 with body composition and muscle structural and functional parameters in a homogenous, healthy, non-obese non-resistance trained population (2) investigate whether *FTO* rs9939609 genotype differed between elite rugby athletes and a control population, and/or between RU player positions. Based on prior data in obese populations, it was hypothesised that the rs9939609 risk (A) allele would be associated with greater body mass, fat mass, BMI, LM, muscle volume and muscle strength in non-resistance trained individuals. Secondly, for the elite rugby cohort, it was hypothesised that the *FTO* A allele would be overrepresented in player positions typically requiring greater body and muscle mass while the protective (T) allele would be more common in positions requiring a lean phenotype.

Method

Participants

A total of 1089 individuals were recruited and gave written informed consent to participate in the present study. The total sample comprised elite Caucasian male rugby athletes ($n = 530$; height 1.85 (0.07) m, mass 101 (14) kg, age 29 (7) yr, BMI 29.4 (3.7) kg·m^{-2}; mean (standard deviation (SD)) including 73% British, 16% South African, 7% Irish and 4% from other nationalities and non-athlete Caucasian control participants (male and female; $n = 559$; height 1.75 (0.10) m, mass 75 (13) kg, age 29 (16) yr, BMI 24.5 (3.6) kg·m^{-2}) including 86% British, 12% South African, 1% Irish and 1% from other nationalities. Athletes were considered elite if they had competed regularly (> 5 matches) since 1995 in the highest professional league in the UK, Ireland or South Africa for RU or the highest professional league in the UK for rugby league (RL). Of the RU athletes, 52.7% had competed at an international level for a "High Performance Union" (Regulation 16, worldrugby.org) and 43.2% of RL had competed at international level. All data for the athlete group's international status were confirmed as of 1st June 2016. Furthermore, within the rugby cohort, a subsample ($n = 77$) were examined for performance-related muscle phenotypes. Within the control group were a subgroup of non-resistance trained healthy males (NT; Table 1). NT participants were aged 18–39 years, had a BMI 18.5–30 kg·m^{-2}, had not undertaken any structured resistance training in the preceding 12 months and had no history of neurological or musculoskeletal disorders. Additionally, only those participants undertaking less than 3 h of low-to-moderate physical activity per week, assessed via questionnaire [24], were included.

Table 1 Descriptive, morphological and functional characteristics of all participants and genetic frequency, in the non-resistance trained (NT) cohort

Phenotype	All (*n* = 120)	AA (*n* = 18)	AT (*n* = 58)	TT (*n* = 44)	TT + AT (*n* = 102)	*P* value Additive (Recessive)
Height (m)	1.79 (0.06)	1.80 (0.06)	1.78 (0.07)	1.80 (0.06)	1.79 (0.07)	0.15 (0.46)
Mass (kg)	75.0 (10.0)	81.0 (8.1)	74.2 (10.5)	73.5 (9.6)	73.9 (10.0)	0.03 (0.02)
BMI (kg m^{-2})	23.4 (2.7)	25.1 (2.6)	23.5 (2.8)	22.6 (2.4)	23.1 (2.6)	0.02 (0.02)
Age (years)	20.6 (2.3)	21.6 (2.8)	20.9 (2.4)	19.7 (1.6)	20.4 (2.1)	0.02 (0.08)
Fat mass (%)	21.5 (5.2)	23.3 (5.5)	21.8 (5.2)	20.3 (4.8)	21.2 (5.1)	0.09 (0.11)
LM (%)	73.5 (5.9)	70.3 (6.3)	73.3 (5.6)	75.1 (5.8)	74.4 (5.6)	0.04 (0.04)
Total appendicular LM (%)	33.4 (3.8)	31.6 (3.5)	33.0 (3.7)	34.6 (3.7)	33.7 (3.7)	0.04 (0.05)
Arm LM (%)	8.5 (1.2)	7.9 (1.0)	8.6 (1.3)	8.7 (1.1)	8.7 (1.2)	0.06 (0.04)
Leg LM (%)	24.8 (2.9)	23.7 (3.1)	24.4 (2.8)	25.8 (2.9)	25.0 (2.9)	0.04 (0.10)
V_{VL} (cm^3)	566 (86)	585 (81)	550 (86)	580 (85)	563 (86)	0.20 (0.61)
VL ACSA (cm^2)	21.4 (2.5)	21.9 (2.8)	21.0 (2.4)	21.9 (2.5)	21.4 (2.5)	0.20 (0.86)
VL PCSA (cm^2)	71.7 (13.9)	71.2 (13.9)	72.0 (13.7)	71.5 (11.0)	71.8 (12.6)	0.97 (0.61)
Isometric MVC_{KE} torque (N m)	272 (53)	285 (38)	271 (63)	268 (45)	270 (56)	0.50 (0.25)
VL specific force (Ncm^{-2})	21.6 (2.6)	22.3 (2.5)	21.6 (2.5)	21.1 (2.8)	21.4 (2.6)	0.50 (0.25)

Data are mean (SD)
BMI body mass index, *LM* lean mass, *VL* vastus lateralis, *ACSA* anatomical cross sectional area, *PCSA* physiological cross-sectional area, MVC_{KE} maximal voluntary contraction, V_{VL} vastus lateralis muscle volume

Procedures

Muscle properties in NT

An isokinetic dynamometer (Cybex Norm, Cybex International Inc., NY, USA) was used to assess maximal isometric knee extension (MVC_{KE}) and maximal isometric knee flexion (MVC_{KF}) torque at knee joint angles of 70°, 80° and 90° (full knee extension = 0°). The angle of peak torque was taken as the optimal knee joint angle and used to estimate antagonist muscle co-activation during MVC, which assumed a linear relationship between biceps femoris EMG activity and knee flexion torque [25]. Together with antagonist co-activation, quadriceps femoris (QF) voluntary activation capacity, determined using the interpolated twitch technique [26], allowed for the calculation of net MVC_{KE} torque. Subsequently, patella tendon moment arm length (d_{PT}) was measured using dual energy x-ray absorptiometry DXA; [27] and patella tendon force calculated as net MVC_{KE} torque/d_{PT}. The contribution of the vastus lateralis (VL) muscle (MF_{VL}) to patella tendon force was calculated by estimating the relative physiological cross-sectional area (PCSA) of the VL as ~21% of the QF [28]. VL muscle architecture (fascicle length, L_f and pennation angle, θ) was measured at 50% of VL length during MVC_{KE} at the optimal knee joint angle using ultrasound (AU5, Esaote, Italy) and VL fascicle force estimated as MF_{VL}/cosine θ [29]. At rest, ultrasound was also used to obtain a series of transverse plane scans at 50% of VL muscle length from the medial to lateral borders, which were contour matched to determine VL anatomical cross-sectional area [ACSA; 29]. With VL length and VL ACSA used to estimate VL volume (V_{VL}) as previously [30]. Subsequently, VL PCSA was calculated as V_{VL}/L_f and, VL specific force calculated by dividing VL fascicle force by VL PCSA [29]. Finally, quantification of whole body and appendicular LM was completed using DXA (Hologic Discovery: Vertec Scientific Ltd, UK) following a 12 h overnight fast. Participants lay in a supine position, avoiding any contact between the trunk and the appendicular mass during a 7 min whole body scanning procedure (effective dose, 8.4 μSv). Appendicular lean mass was estimated from the DEXA as the total muscle mass of both the upper and lower limbs, where LM is body mass excluding fat and bone mass.

Muscle power in RU athletes

Using a portable force platform (Type 92866AA, Kistler, Germany) peak power output (PPO) was determined during a bilateral countermovement jump (CMJ) according to methods described previously [31]. Body mass and the vertical component of the ground reaction force during the CMJ (sampled at 1000 Hz) were used to determine instantaneous velocity and displacement of the participant's centre of gravity. Instantaneous power output was determined using the following equation: Power (W) = vertical GRF (N) x vertical velocity of centre of gravity ($m \cdot s^{-1}$), with the highest value produced deemed the PPO.

Sample collection and genotyping

Description of all molecular procedures have previously been described in detail [22]. Briefly, blood (~70% of all

samples), saliva (~25%) or buccal swab samples (~5%) were obtained via the following protocols. Blood was drawn from a superficial forearm vein into an EDTA tube and stored in sterile tubes at -20°C until processing. Saliva samples were collected into Oragene DNA OG-500 collection tubes (DNA Genotek Inc., Ontario, Canada) according to the manufacturer's protocol and stored at room temperature until processing. Sterile buccal swabs (Omni swab, Whatman, Springfield Mill, UK) were rubbed against the buccal mucosa of the cheek for approximately 30 s. Tips were ejected into sterile tubes and stored at -20°C until processing. DNA isolation and genotyping were performed in the MMU, University of Glasgow, University of Cape Town (DNA isolation only) and University of Northampton laboratories. The majority of samples were processed and genotyped in the MMU laboratory, including all samples within NT. At MMU and Glasgow, DNA isolation was performed using the QIAamp DNA Blood Mini kit and standard spin column protocol, following the manufacturer's instructions (Qiagen, West Sussex, UK). Briefly, 200 *μl*. of whole blood/saliva, or one buccal swab, was lysed, incubated, the DNA washed and the eluate containing isolated DNA stored at 4°C. In Cape Town, DNA was isolated from whole blood using a different protocol [32] during which samples were lysed, centrifuged, the DNA washed and samples stored at -20°C. Genotyping of DNA isolated in Cape Town was performed in Glasgow. At Northampton, DNA was isolated from whole blood using Flexigene kits (Qiagen). Briefly, samples were lysed, DNA precipitated and washed, with samples stored at -20°C.

Genotyping in all three genotyping centres was performed on *FTO* (*rs9939609*). Briefly, in the Glasgow laboratory 10 *μl*. Genotyping Master Mix (Applied Biosystems, Paisley, UK), 1 *μl*. SNP-specific TaqMan assay (Applied Biosystems), 6 *μl*. nuclease-free H_2O and 3 *μl*. DNA solution (~9 ng DNA) were added per well. In the Northampton laboratory, genotyping was performed by combining 10 *μl*. of Genotyping Master Mix, 8 *μl*. H_2O, 1 *μl*. assay mix with 1 *μl*. of purified DNA (~10 ng). In both laboratories, PCR was performed using a StepOnePlus real-time detector (Applied Biosystems). Briefly, after an initial 10 min at 95°C, 40 cycles of denaturation at 92°C for 15 s then annealing and extension at 60°C for 1 min were used. Genotyping calls were performed using StepOnePlus software version 2.3 (Applied Biosystems). At MMU, 5 *μl*. Genotyping Master Mix, 4.3 *μl*. H_2O, 0.5 *μl*. assay mix and 0.2 *μl*. of purified DNA (~9 ng) were used in each reaction for samples derived from blood and saliva. For DNA derived from buccal swabs, 5 *μl*. Genotyping Master Mix was combined with 3.5 *μl*. H_2O, 0.5 *μl*. assay mix and 1 *μl*. DNA solution (~9 ng DNA). Either a Chromo4 (Bio-Rad, Hertfordshire, UK) or StepOnePlus real-time PCR system was used. Briefly, after an initial 10 min at 95°C, 40 cycles of denaturation at 92°C for 15 s then annealing and extension at 60°C for 1 min were used. Genotyping calls were performed using Opticon Monitor software version 3.1 (Bio-Rad) or StepOnePlus software version 2.3. The Taqman assay included VIC and FAM dyes that indicated A and T alleles on the forward DNA strand, respectively. Thus, VIC/FAM were interpreted as: 5'-GTGAATTT[A/T]GTGATGCA-3'.

RU positional groups

As established previously [22], to compare genotype and allele frequencies within the RU group, athletes were allocated to subgroups; forwards (props, hookers, locks, flankers, number eights) and backs (scrum halves, fly halves, centres, wings, full backs). Also, due to diverse physiological demands within RU [33, 34], athletes were further divided into established positional groups according to their movement patterns [33]; front five (props, hookers, locks), back row (flankers, number eights), half backs (scrum halves, fly halves), back three (wings and full backs) and centres. For example, in one study the front five travelled ~136 m at >5 $m \cdot s^{-1}$ compared to ~566 m for the back three [35]. Comparisons between positions were not performed for the RL cohort due to low statistical power when it was subdivided.

Data analysis

SPSS for Windows version 22 (SPSS Inc., Chicago, IL) software was used to conduct statistical analyses. One-way analysis of variance (ANOVA) was used to compare height, mass, BMI, age and PPO between sample populations and genotype groups. When genotype groups were compared using a recessive model, an independent samples t-test was used. Furthermore, genotype effects on muscle phenotypes of interest were assessed for linear trend. Pearson's χ^2 tests compared genotype and allelic frequencies between athlete and control groups and between RU positional subgroups. There were 30 comparisons for genotype frequency between groups and 28 tests of genotype differences in phenotype in NT that were subjected to Benjamini-Hochberg corrections to control false discovery rate and corrected probability values are reported [36]. Where appropriate, odds ratio (OR) and eta squared (η^2) were calculated to estimate effect size. Alpha was set at 0.05.

Results

Genotype calling was successful in all samples. There was 100% agreement among reference samples genotyped in the three genotyping centres, i.e. Glasgow, Northampton and MMU laboratories. Genotype frequencies were in Hardy-Weinberg equilibrium for

the entire control group ($P = 0.871$), NT ($P = 0.988$) and athlete groups (RL, $P = 0.183$; RU, $P = 0.076$).

Non-resistance trained (NT)

The AA genotype group had greater body mass, BMI and age but not height compared to other genotype groups (Table 1). There were genotype differences for total body ($\eta_p^2 = 0.072$), total appendicular ($\eta_p^2 = 0.075$) and leg ($\eta_p^2 = 0.078$) LM, with tendencies for arm LM ($P = 0.06$, $\eta_p^2 = 0.054$) and total fat mass ($P = 0.09$, $\eta_p^2 = 0.024$). T-allele carriers demonstrated greater total body (5.3%), appendicular (6.7%) and arm (9.8%), but not leg ($P = 0.10$) LM than AA homozygotes. There were no differences in muscle size, torque or specific force variables (Table 1).

Athletes

Athletes were taller and heavier ($P < 0.05$) but not older ($P > 0.05$) than controls. There were no genotype frequency differences between athletes (RL and RU combined; $P = 0.16$), RL ($P = 0.36$), RU ($P = 0.16$) and controls (only additive models presented).

In terms of player position, backs had a greater frequency of T allele carriers than forwards ($P = 0.03$, Table 2, Fig. 1) and showed greater odds of being T allele carriers than AA genotype (OR = 1.84, Table 3). When combined, the back three and centres group contained less AA homozygotes and more T allele carriers ($P = 0.03$, $P = 0.02$, respectively; Fig. 1a and Table 2) than controls. Additionally, controls had more than twice the odds of being AA than the back three and centres group, with greater odds of T allele carriers in the back three and centres than controls (Table 3). Compared to forwards and all other RU athletes, TT genotype ($P = 0.03$; $P = 0.03$, respectively) and T allele carriers ($P = 0.02$; $P = 0.02$, respectively) were more common in the back three and centres group (Fig. 1a and Table 2). Likewise, forwards and all other RU athletes had greater than three times the odds of being AA genotype than the back three and centres group, with greater odds of T allele carriers in the back three and centres group than forwards and all other RU athletes (Table 3). Furthermore, the back three and centres group showed a greater T allele frequency than both forwards and all other RU athletes (Fig. 1b) and almost one and a half times greater odds of possessing the T allele (Table 3).

Muscle power

While PPO tended to be greater in back row players with lowest power in the halfbacks (5792 vs 5000 W; $P = 0.09$), PPO relative to body mass did differ according to playing position (Table 4). The centre and back 3 players (59 Wkg^{-1}) were 9.8% more powerful than back row (54 Wkg^{-1}; $P = 0.025$) and 20.2% more powerful than front 5 players (47 Wkg^{-1}; $P = 6 \times 10^{-7}$; Table 4).

Discussion

We have shown that individuals possessing the *FTO* rs9939609 T allele and TT genotype had greater LM, while no differences in leg muscle size or strength were observed (Table 1), thus rejecting our first hypothesis that the risk (A) allele would be associated with greater LM and muscle volume in a healthy non-resistance trained population. That greater LM in T allele and TT genotype individuals was observed despite A allele carriers having greater body mass and BMI as reported previously [5, 6]. In agreement with our second hypothesis, we report a greater T allele and TT genotype frequency in elite rugby athlete playing positions more reliant on a lean phenotype for success [37], while the A allele is more common in those positions where total body mass is more important [38]; Fig. 1, Tables 2 and 3. The ability to rapidly produce high levels of power relative to body mass using the leg musculature was greater in those playing positions more reliant on a lean phenotype (Table 4) and is in agreement with previously published data of elite RU players [39]. One possible biological mechanism underlying the present results may be the action of the iroquois homeobox 3 (IRX3) protein through its *FTO* genomic loci interaction.

Table 2 Genotype and allele distribution of controls and athletes separated by code (RL and RU) and into positional subgroups for RU, presented as genotype/allele counts followed by percentage in parentheses

Genotype/allele	All athletes	RL athletes	RU athletes	Controls	Forwards	Backs	Front 5	Back row	Half backs	Back 3 and centres
FTO										
AA	80 (15.1)	12 (13.6)	69 (15.3)	90 (16.1)	48 (18.5)	21 (11.0)	30 (17.0)	18 (21.7)	13 (17.8)	8 (6.8)*
AT	280 (52.7)	49 (55.7)	235 (52.3)	266 (47.6)	133 (51.4)	102 (54.7)	94 (53.4)	39 (47.0)	34 (46.6)	68 (57.6)
TT	170 (32.2)	27 (30.7)	146 (32.4)	203 (36.3)	78 (30.1)	68 (34.3)	52 (29.6)	26 (31.3)	26 (35.6)	42 (35.6)*
Total	530	88	450	559	259	191	176	83	73	118
A allele	440 (41.5)	73 (41.5)	375 (41.4)	446 (39.9)	229 (44.2)	144 (37.7)	154 (43.7)	75 (45.2)	60 (41.1)	84 (35.6)*
T allele	620 (58.5)	103 (58.5)	527 (58.6)	672 (60.1)	289 (55.8)	238 (62.3)	198 (56.3)	91 (54.8)	86 (58.9)	152 (64.4)

Eight athletes competed in both elite RL and RU and were included in both groups that were analysed separately. *Different from controls ($P < 0.06$). *Different from forwards ($P = 0.03$)

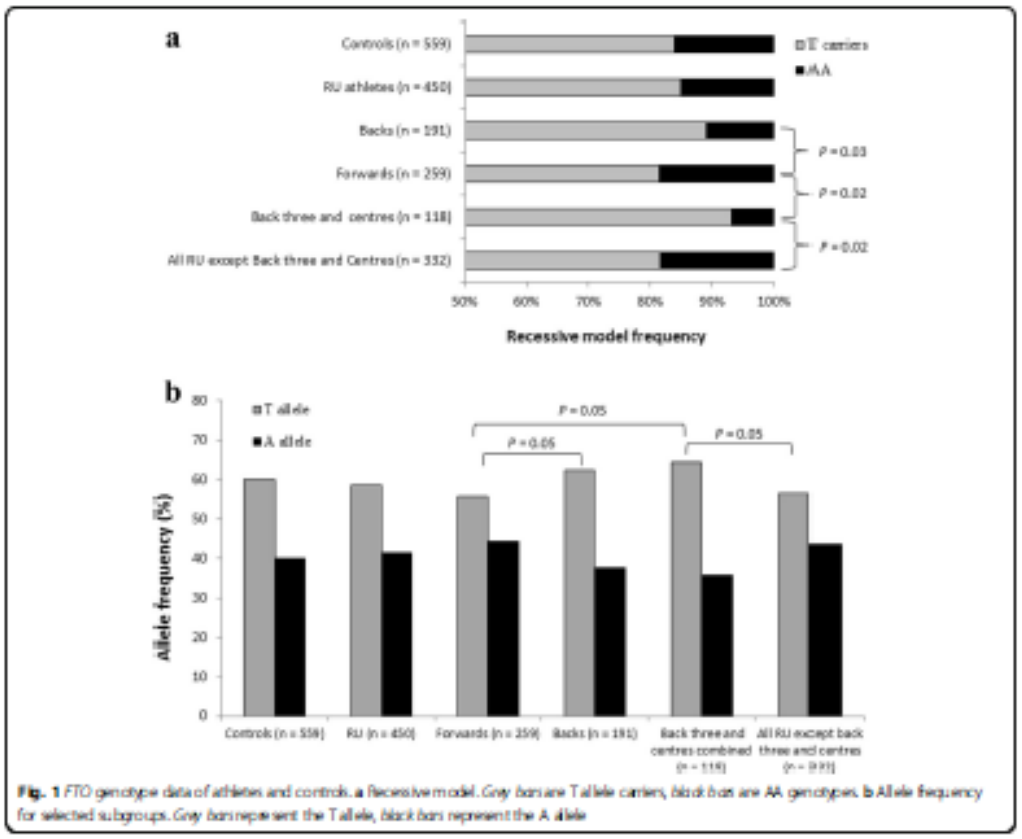


Fig. 1 *FTO* genotype data of athletes and controls. **a** Recessive model. *Gray bars* are T allele carriers, *black bars* are AA genotypes. **b** Allele frequency for selected subgroups. *Gray bars* represent the T allele, *black bars* represent the A allele

Table 3 Odds Ratio statistics for RU athlete status by playing position for *FTO* genotype (TT/AA), allele (T/A) and recessive (T/AA) genetic models

Positional comparison	Genetic model	Odds ratio	95% Confidence Interval	*P* value
Backs v Forwards	T/AA	1.84	1.06-3.19	0.029
Back 3 and centres v Controls	TT/AA	2.33	1.05-5.16	0.038
	T/AA	2.64	1.05-5.16	0.012
Back 3 and centres v Forwards	TT/AA	3.23	1.39-7.46	0.006
	T/AA	3.12	1.43-6.84	0.004
	T/A	1.44	1.04-1.97	0.026
Back 3 and centres v all other RU athletes	TT/AA	3.08	1.36-6.98	0.007
	T/AA	3.09	1.43-6.68	0.004
	T/A	1.37	1.01-1.86	0.045
Back 3 and centres v other backs	TT/AA	2.98	1.17-7.59	0.022
	T/AA	2.63	0.96-7.19	0.060

Until recently, little was known about the molecular basis for *FTO* SNP associations with any reported phenotype measure, because there was no association between *FTO* SNPs and expression of the *FTO* protein [40, 41]. However, *FTO* has recently been found to influence IRX3 protein expression, through evolutionarily conserved long-range chromatin looping. Individuals possessing the protective *FTO* genotype/allele (TT/T) display lower IRX3 expression than AA homozygotes [42]. Furthermore, in contrast to IRX3 knockout (KO) mice, wild type mice exhibited similar *FTO* SNP risk (A) allele-associated phenotypes, such as greater BMI, body mass and body fat percentage [42]. Interestingly, IRX3 KO mice expended more energy, particularly at night, due to a greater browning of white fatty tissue [42] and recent findings show a link between brown fat and muscle developmental precursor Myf5 [43] which may provide a mechanism for the observation of greater LM in *FTO* T allele carriers in our NT cohort. Moreover,

Table 4 Muscle power of RU athletes ($n = 77$) in positional groups. P values are from comparisons of power between the four groups. Data are mean (SD)

Phenotype	Front 5 ($n = 32$)	Back row ($n = 14$)	Half backs ($n = 14$)	Back 3 and centres ($n = 17$)	P value
Power (W)	5592 (819)	5687 (858)	4937 (650)	5579 (569)	0.030
Relative Power ($W\,kg^{-1}$)	49.4 (7.7)	52.7 (6.9)	56.1 (7.2)	59.8 (4.2)*	2×10^{-9}

*Different from all other players ($P = 8 \times 10^{-5}$), including front 5 ($P = 3 \times 10^{-8}$) and back row ($P = 0.005$)

using a transgenic mouse model ($Rosa26^{Tav\text{-}Irx3}$) that disrupts IRX3 function whilst maintaining the genomic interaction between IRX3 and *FTO* (mimicking more accurately the human in vivo state than the aforementioned KO model), the authors showed retention of the KO model phenotype traits [42]. These *FTO*-IRX3 protein interactions suggest an explanation for the greater LM seen with the T allele carriers of the present study and possibly the association of the T allele with muscle power relative to body mass and its relationship with playing position in RU athletes (Tables 1, 2, and 4; Fig. 1).

The precise mechanisms of action of IRX3 in mammalian physiology are not fully understood, however the primary role of IRX3 in embryonic development and future actions in motor neuron restriction is relevant to this discussion. During neuronal development, IRX3 expression plays a key role in N-tubulin development and initiation of neuronal programming. High levels of IRX3 protein promote early tissue development resulting in a lack of cell differentiation [44]. Thus, it is possible that because the *FTO* T allele is associated with lower IRX3 expression, greater early motor neuron differentiation might subsequently lead to greater LM – as we observed (Table 1). As such, for predeterminant neuronal cells to differentiate into a progenitor motor neuron domain and subsequently motor neurons, it appears IRX3 must be repressed by the microRNA *miR-17-3p* in order for OLIG2 to regulate the development of ventral spinal motor neurons [45]. Thus, as the expression of OLIG2 increases, the yield of motor neurons increases in tandem [46]. Considering *FTO* T allele carriers have a lower embryonic expression of IRX3, T allele carriers may have a predisposition for greater LM through enhanced life-long motor neuron availability via OLIG2 expression and therefore, may be at an advantage for certain forms of athletic ability and associated performance phenotypes (Tables 2, 3 and 4; Fig. 1). This rationale and the present results may represent a small portion of the 85% heritability of adult muscle neuronal function [47].

Recent associations between *FTO* variants and IGF-1, specifically that serum IGF-1 levels were greater in T allele carriers [48], may provide a second mechanism to explain the observed genotype differences in LM (Table 1). It is well known that IGF-1 is upregulated as a consequence of mechanical load/exercise and plays an important role in the cellular development of muscle hypertrophy [49]. Hence, T allele carriers, who in the NT group had significantly greater LM, may experience upregulation of IGF-1 compared to AA genotype counterparts. Furthermore, serum IGF-1 levels have been positively associated with quadriceps torque [50] and explosive muscle power [51] in older men. These data provide a further potential basis for our observation that RU athletes who require greater muscle power relative to body mass (Table 4) show a greater frequency of the T allele than other playing positions (Table 2 and 3; Fig. 1).

We observed a lower frequency of the AA genotype in back three and centre playing positions (OR = 2.53; Table 3), although there was no difference between the entire rugby cohort and controls. That latter observation is similar to that we reported previously regarding another genetic variant (*ACTN3* rs1815739) where there was no difference between the entire rugby cohort and controls despite differences in genotype frequency according to playing position [22]. This again demonstrates the importance of defining athletes very carefully when conducting such comparisons. Global positioning system (GPS) data provide evidence for the relevance of our finding regarding *FTO* genotype in elite athletes. Jones et al. [35] showed that - at an elite competitive level - the back three and centre players express the greatest 'instantaneous and accumulative demands for exercise' (exertion index; EI) than all other athletes and spent more time at sprinting intensities. Thus, there is congruence between our dual observations of firstly greater LM in NT associated with the T allele, and secondly a greater frequency of the T allele in certain elite rugby athletes who rely on greater power outputs relative to body mass to be successful in those specific playing positions Table 3; [39].

These data suggest the relevance of the *FTO* rs9939609 T allele to muscle-related phenotypes and subsequently, athletic success. When considering the possible molecular mechanism from *FTO* via IRX3 to OLIG2 resulting in greater lifelong motor neuron availability, this may have implications for muscle size-related disorders such as sarcopenia and cachexia.

Conclusions

The presented data show a novel dimension of *FTO* genetic variation in human physiology, by investigating in vivo muscle phenotypes in a healthy non-resistance trained population and relating those data to the extreme upper

end of human physical performance – i.e. elite athletes. We show that the *FTO* rs9939609 protective T allele may be responsible for part of the inherited component of the inter-individual variability in LM. Furthermore, elite athletes who rely greatly on LM relative to total body mass for athletic success (RU back three and centre players, in this study) also seem more likely to carry a protective T allele, have higher peak muscle power output relative to body mass and are likely to selected for appropriate playing positions as a result of these and other phenotypes. It is possible that this association between the *FTO* rs9939609 SNP, via LM, and athletic success, could be a result of the interaction with IRX3 in embryonic development of motor neuron patterning and the IGF-1 muscle development pathway. The strengths of the presented paper are the two-layered study design (NT and elite athlete cohorts) and the combination of muscle functional phenotypes with case-control data. While the finding that *FTO* genotype differs among elite rugby playing positions is a new insight, *FTO* is only one of many variants (most others unknown) that contribute to this phenotype and as such should not be used for talent identification at this time. Replication is necessary for each cohort using comparable methods, and future experimental focus should be on the proposed biological pathways of these *FTO* associations with muscle phenotypes.

Abbreviations

ACSA: Anatomical cross-sectional area; BMI: Body mass index; CMJ: Countermovement jump; DXA: Dual energy x-ray absorptiometry; FTO: Fat mass and obesity; IGF-1: Insulin-like growth factor 1; IRX3: Iroquois homeobox 3; LM: Lean mass; MVC_{EE}: Maximal isometric knee extension; MVC_{EF}: Maximal isometric knee flexion; NT: Non-resistance trained; OLIG2: Oligodendrocyte transcription factor; PCSA: Physiological cross-sectional area; PPO: Peak power output; QF: Quadriceps femoris; RL: Rugby league; RU: Rugby union; VL: Vastus lateralis; V_M: VL muscle volume

Acknowledgements

The authors wish to thank all athletes, their respective scientific support staff and the control participants for their time and willingness to participate in this research. We would like to thank Louis B. Khoury for his technical assistance at the University of Northampton and Ben Davies for assistance in participant recruitment.

Funding

Research funded by Manchester Metropolitan University.

Availability of data and materials

Data used to generate statistical conclusions are presented in the published article. Researchers may reasonably request additional data and materials from the RugbyGene project, via contacting the corresponding author.

Authors' contributions

Listed alphabetically: MC, SD, RE, SH, IK, YP, AW and GW conceived and designed the study: MB, CC, MC, RE, SH, LK, CM, BM, CR, SR, WR, GS, BV, AW and GW contributed to data collection. SH and AW analysed data and drafted the manuscript. All authors contributed to interpretation of data, revised the article critically for important intellectual content and approved the final version of the manuscript.

Competing interests

The authors declare that they have no competing interests.

Consent for publication

Not applicable.

Ethics approval and consent to participate

Ethical approval was granted by the ethics committees of Manchester Metropolitan University, Glasgow University, University of Cape Town and University of Northampton and all experimental procedures complied with the Declaration of Helsinki (2013). All participants gave written informed consent to take part in the present study and for the results to be published.

Twitter

Follow the RugbyGene project at @RugbyGeneStudy

Author details

[1]MMU Sports Genomics Laboratory, Manchester Metropolitan University, Crewe Green Road, Crewe CW1 5DU, UK. [2]A-STEM, College of Engineering, Swansea University, Swansea, UK. [3]Research Institute for Sport & Exercise Sciences, Liverpool John Moores University, Liverpool, UK. [4]Institute of Sport, Exercise and Health, University College London, London, UK. [5]Centre for Physical Activity and Chronic Disease, Institute of Health and Wellbeing, University of Northampton, Northampton, UK. [6]School of Sport, Health and Exercise Sciences, Bangor University, Bangor, UK. [7]Institute of Cardiovascular & Medical Sciences University of Glasgow, Glasgow, UK. [8]Medical and Scientific Department, South African Rugby Union, Cape Town, South Africa. [9]Discipline of Sports Science, Faculty of Health Sciences, University of Kwazulu-Natal, Durban, South Africa. [10]IOC Research Collaborating Centre of Sports Medicine for Anti-Doping Research, University of Brighton, Brighton, UK. [11]Division of Exercise Science and Sports Medicine, Department of Human Biology, University of Cape Town (UCT), Cape Town, South Africa. [12]School of Healthcare Science, Manchester Metropolitan University, Manchester, UK.

Received: 18 August 2016 Accepted: 10 January 2017
Published online: 19 January 2017

References

1. Jacobson JA, Schilsth H, Fredrikson R. The impact of intronic single nucleotide polymorphisms and ethnic diversity for studies on the obesity gene FTO. Obes Rev. 2012;13(12):1085–109.
2. He Q, Fu M, Mao S, Hatta K, Chandak GR, Xi B. FTO gene variant and risk of hypertension: A meta-analysis of 57,464 hypertensive cases and 41,256 controls. Metabolism. 2014;63(5):633–9.
3. Liu C, Mou S, Pan C. The FTO gene rs9939609 polymorphism predicts risk of cardiovascular disease: a systematic review and meta-analysis. PLoS ONE. 2013;8(8):e71901.
4. Loos RJ, Yeo GS. The bigger picture of FTO: the first GWAS-identified obesity gene. Nat Rev Endocr. 2014;10(1):51–61.
5. Frayling TM, Timpson NJ, Weedon MN, Zeggini E, Freathy RM, Lindgren CM, et al. A common variant in the FTO gene is associated with body mass index and predisposes to childhood and adult obesity. Science. 2007; 316(5826):889–94.
6. Rauhio A, Uusi-Rasi K, Nikkari ST, Kannus P, Sievänen H, Kunnas T. Association of the FTO and ADRB2 genes with body composition and fat distribution in obese women. Maturitas. 2013;76(2):165–71.
7. Andreasen CH, Stender-Petersen KL, Mogensen MS, Torekov SS, Wegner L, Andersen G, et al. Low physical activity accentuates the effect of the FTO rs9939609 polymorphism on body fat accumulation. Diabetes. 2008;57(1):95–101.
8. Liu G, Zhu H, Lagou V, Guth B, Stallmann-Jorgensen IS, Treiber FA, et al. FTO variant rs9939609 is associated with body mass index and waist circumference, but not with energy intake or physical activity in European- and African-American youth. BMC Med Genet. 2010;11(1):57.
9. Sonestedt E, Gullberg B, Ericson U, Wirfält E, Hedblad B, Orho-Melander M. Association between fat intake, physical activity and mortality depending on genetic variation in FTO. Int J Obesity. 2011;35(8):1041–9.
10. Matsuo T, Nakata Y, Hotta K, Tanaka K. The FTO genotype as a useful predictor of body weight maintenance: Initial data from a 5-year follow-up study. Metabolism. 2014;63(7):912–7.
11. Tanofsky-Kraff M, Han JC, Anandalingam K, Shomaker LB, Columbo KM, Wolkoff LE, et al. The FTO gene rs9939609 obesity-risk allele and loss of control over eating. Am J Clin Nutr. 2009;90(6):1483–8.

12. Jess T, Zimmermann E, Kring SII, Berentzen T, Holst C, Toubro S, et al. Impact on weight dynamics and general growth of the common FTO rs9939609: a longitudinal Danish cohort study. Int J Obesity. 2008; 32(9):1388–94.
13. Livshits G, Malkin I, Moayyeri A, Spector TD, Hammond CJ. Association of FTO gene variants with body composition in UK twins. Ann Hum Genet. 2012;76(5):383–91.
14. Phillips CM, Kesse-Guyot E, McManus R, Hercberg S, Lairon D, Planells R, et al. High Dietary Saturated Fat Intake Accentuates Obesity Risk Associated with the Fat Mass and Obesity–Associated Gene in Adults. J Nutr. 2012; 142(5):824–31.
15. Corella D, Arnett DK, Tucker KL, Kabagambe EK, Tsai M, Parnell LD, et al. A high intake of saturated fatty acids strengthens the association between the fat mass and obesity-associated gene and BMI. J Nutr. 2011;141(12):2219–25.
16. Zhang X, Qi Q, Zhang C, Smith SR, Hu FB, Sacks FM, et al. FTO Genotype and 2-Year Change in Body Composition and Fat Distribution in Response to Weight-Loss Diets: The POUNDS LOST Trial. Diabetes. 2012;61(11):3005–11.
17. Huang T, Qi Q, Li Y, Hu FB, Bray GA, Sacks FM, et al. FTO genotype, dietary protein, and change in appetite: the Preventing Overweight Using Novel Dietary Strategies trial. Am J Clin Nutr. 2014;99(5):1126–30.
18. Kilpeläinen TO, Qi L, Brage S, Sharp SJ, Sonestedt E, Demerath E, et al. Physical activity attenuates the influence of FTO variants on obesity risk: a meta-analysis of 218,166 adults and 19,268 children. PLoS Med. 2011;8(11):e1001116.
19. Rankinen T, Rice T, Teran-Garcia M, Rao DC, Bouchard C. FTO Genotype Is Associated With Exercise Training–Induced Changes in Body Composition. Obesity. 2010;18(2):322–6.
20. Mustelin L, Pietiläinen KH, Rissanen A, Naukkarinen J, Sarna S, Kaprio J. Habitually active adolescents are protected against the metabolic consequences of the FTO gene. Int J Obesity. 2009;33(12):1405–11.
21. Mustelin L, Silventoinen K, Pietiläinen K, Rissanen A, Kaprio J. Physical activity reduces the influence of genetic effects on BMI and waist circumference: a study in young adult twins. Int J Obesity. 2008;32(5):873–80.
22. Rankinen T, Sarzynski MA, Ghosh S, Bouchard C. Are there genetic paths common to obesity, cardiovascular disease outcomes, and cardiovascular risk factors? Circ Res. 2015;116(5):909–22.
23. Hautala AJ, Kiviniemi AM, Mäkikallio TH, Kinnunen H, Nissilä S, Huikuri HV, et al. Individual differences in the responses to endurance exercise training and the role of the insulin sensitivity gene IRS1. Eur J Appl Physiol. 2010;108(3):511–9.
24. Hautala AJ, Kiviniemi AM, Mäkikallio TH, Kinnunen H, Nissilä S, Huikuri HV, et al. Individual differences in the responses to endurance exercise training and the role of the insulin sensitivity gene IRS1. Eur J Appl Physiol. 2010;108(3):511–9.
25. Rankinen T, Sarzynski MA, Ghosh S, Bouchard C. Are there genetic paths common to obesity, cardiovascular disease outcomes, and cardiovascular risk factors? Circ Res. 2015;116(5):909–22.
26. Mustelin L, Silventoinen K, Pietiläinen K, Rissanen A, Kaprio J. Physical activity reduces the influence of genetic effects on BMI and waist circumference: a study in young adult twins. Int J Obesity. 2008;32(5):873–80.
27. Sarzynski MA, Rankinen T, Kraus WE, Church TS, Bouchard C. Associations of Physical Activity with Insulin Sensitivity and Glucose Tolerance. Med Sci Sports Exerc. 2015;47(8):1589–96.
28. Hautala AJ, Kiviniemi AM, Mäkikallio TH, Kinnunen H, Nissilä S, Huikuri HV, et al. Individual differences in the responses to endurance exercise training and the role of the insulin sensitivity gene IRS1. Eur J Appl Physiol. 2010;108(3):511–9.
29. Rankinen T, Sarzynski MA, Ghosh S, Bouchard C. Are there genetic paths common to obesity, cardiovascular disease outcomes, and cardiovascular risk factors? Circ Res. 2015;116(5):909–22.
30. Mustelin L, Silventoinen K, Pietiläinen K, Rissanen A, Kaprio J. Physical activity reduces the influence of genetic effects on BMI and waist circumference: a study in young adult twins. Int J Obesity. 2008;32(5):873–80.
31. Sarzynski MA, Rankinen T, Kraus WE, Church TS, Bouchard C. Associations of Physical Activity with Insulin Sensitivity and Glucose Tolerance. Med Sci Sports Exerc. 2015;47(8):1589–96.
32. Hautala AJ, Kiviniemi AM, Mäkikallio TH, Kinnunen H, Nissilä S, Huikuri HV, et al. Individual differences in the responses to endurance exercise training and the role of the insulin sensitivity gene IRS1. Eur J Appl Physiol. 2010;108(3):511–9.
33. Rankinen T, Sarzynski MA, Ghosh S, Bouchard C. Are there genetic paths common to obesity, cardiovascular disease outcomes, and cardiovascular risk factors? Circ Res. 2015;116(5):909–22.
34. Mustelin L, Silventoinen K, Pietiläinen K, Rissanen A, Kaprio J. Physical activity reduces the influence of genetic effects on BMI and waist circumference: a study in young adult twins. Int J Obesity. 2008;32(5):873–80.
35. Sarzynski MA, Rankinen T, Kraus WE, Church TS, Bouchard C. Associations of Physical Activity with Insulin Sensitivity and Glucose Tolerance. Med Sci Sports Exerc. 2015;47(8):1589–96.
36. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Stat Soc. 1995;57(1):289–300.
37. Smart DJ, Hopkins WG, Gill ND. Differences and changes in the physical characteristics of professional and amateur rugby union players. J Strength Cond Res. 2013;27(11):3033–44.
38. Sedeaud A, Marc A, Schipman J, Tafflet M, Hager JP, Toussaint J-F. How they won Rugby World Cup through height, mass and collective experience. Brit J Sports Med. 2012;46:580–4.
39. Crewther B, Kilduff L, Cook C, Cunningham D, Bunce P, Bracken R, et al. Scaling strength and power for body mass differences in rugby union players. J Sports Med Phys Fitness. 2012;52(1):27–32.
40. Grunnet LG, Nilsson E, Ling C, Hansen T, Pedersen O, Groop L, et al. Regulation and function of FTO mRNA expression in human skeletal muscle and subcutaneous adipose tissue. Diabetes. 2009;58(10):2400–8.
41. Wåhlén K, Sjölin E, Hoffstedt J. The common rs9939609 gene variant of the fat mass-and obesity-associated gene FTO is related to fat cell lipolysis. J Lipid Res. 2008;49(3):607–11.
42. Smemo S, Tena JJ, Kim K-H, Gamazon ER, Sakabe NJ, Gómez-Marín C, et al. Obesity-associated variants within FTO form long-range functional connections with IRX3. Nature. 2014;507(7492):371–5.
43. Schulz TJ, Huang TL, Tran TT, Zhang H, Townsend KL, Shadrach JL, et al. Identification of inducible brown adipocyte progenitors residing in skeletal muscle and white fat. P Natl A Sci. 2011;108(1):143–8.
44. Bellefroid EJ, Kobbe A, Gruss P, Pieler T, Gurdon JB, Papalopulu N. Xiro3 encodes a Xenopus homolog of the Drosophila Iroquois genes and functions in neural specification. EMBO J. 1998;17(1):191–203.
45. Chen J-A, Huang Y-P, Mazzoni EO, Tan GC, Zavadil J, Wichterle H. Mir-17-3p controls spinal neural progenitor patterning by regulating Olig2/Irx3 cross-repressive loop. Neuron. 2011;69(4):721–35.
46. Lamas NJ, Johnson-Kerner B, Roybon L, Kim YA, Garcia-Diaz A, Wichterle H, et al. Neurotrophic requirements of human motor neurons defined using amplified and purified stem cell-derived cultures. PLoS ONE. 2014;9(10):e110304.
47. Mero T, Gataullin N, Khasanov V. Genetic variation of maximal velocity and EMG activity. Int J Sports Med. 2003;29(3):177–81.
48. Rankinen T, Sarzynski MA, Ghosh S, Bouchard C. Are there genetic paths common to obesity, cardiovascular disease outcomes, and cardiovascular risk factors? Circ Res. 2015;116(5):909–22.
49. Mustelin L, Silventoinen K, Pietiläinen K, Rissanen A, Kaprio J. Physical activity reduces the influence of genetic effects on BMI and waist circumference: a study in young adult twins. Int J Obesity. 2008;32(5):873–80.
50. Sarzynski MA, Rankinen T, Kraus WE, Church TS, Bouchard C. Associations of Physical Activity with Insulin Sensitivity and Glucose Tolerance. Med Sci Sports Exerc. 2015;47(8):1589–96.
51. Hautala AJ, Kiviniemi AM, Mäkikallio TH, Kinnunen H, Nissilä S, Huikuri HV, et al. Individual differences in the responses to endurance exercise training and the role of the insulin sensitivity gene IRS1. Eur J Appl Physiol. 2010;108(3):511–9.

Physiol Genomics 48: 196–201, 2016.
First published January 12, 2016; doi:10.1152/physiolgenomics.00107.2015.

Association of *ACTN3 R577X* but not *ACE* I/D gene variants with elite rugby union player status and playing position

S. M. Heffernan,[1] L. P. Kilduff,[2] R. M. Erskine,[3,4] S. H. Day,[1] J. S. McPhee,[5] G. E. McMahon,[1,6] G. K. Stebbings,[1] J. P. H. Neale,[1] S. J. Lockey,[1] W. J. Ribbans,[7] C. J. Cook,[8] B. Vance,[9] S. M. Raleigh,[7] C. Roberts,[10,11] M. A. Bennett,[2] G. Wang,[12] M. Collins,[13] Y. P. Pitsiladis,[12] and A. G. Williams[1,4]

[1]*MMU Sports Genomics Laboratory, Manchester Metropolitan University, Crewe, United Kingdom;* [2]*A-STEM, College of Engineering, Swansea University, Swansea, United Kingdom;* [3]*Research Institute for Sport & Exercise Sciences, Liverpool John Moores University, Liverpool, United Kingdom;* [4]*Institute of Sport, Exercise and Health, University College London, London, United Kingdom;* [5]*School of Healthcare Science, Manchester Metropolitan University, Manchester, United Kingdom;* [6]*Northern Ireland Sports Institute, Newtownabbey, Belfast, United Kingdom;* [7]*Division of Sport, Exercise and Life Science, University of Northampton, Northampton, United Kingdom;* [8]*School of Sport, Health and Exercise Sciences, Bangor University, Bangor, United Kingdom;* [9]*Institute of Cardiovascular & Medical Sciences University of Glasgow, Glasgow, United Kingdom;* [10]*Medical and Scientific Department, South African Rugby Union, Cape Town, South Africa;* [11]*Discipline of Sports Science, Faculty of Health Sciences, University of Kwazulu-Natal, Durban, South Africa;* [12]*Centre for Sport and Exercise Science and Medicine (SESAME), University of Brighton, Brighton, United Kingdom; and* [13]*MRC/UCT Research Unit for Exercise Science and Sports Medicine, University of Cape Town (UCT), Cape Town, South Africa*

Submitted 27 October 2015; accepted in final form 7 January 2016

Heffernan SM, Kilduff LP, Erskine RM, Day SH, McPhee JS, McMahon GE, Stebbings GK, Neale JPH, Lockey SJ, Ribbans WJ, Cook CJ, Vance B, Raleigh SM, Roberts C, Bennett MA, Wang G, Collins M, Pitsiladis YP, Williams AG. Association of *ACTN3 R577X* but not *ACE* I/D gene variants with elite rugby union player status and playing position. *Physiol Genomics* 48: 196–201, 2016. First published January 12, 2016; doi:10.1152/physiolgenomics.00107.2015.—We aimed to quantify the *ACE* I/D and *ACTN3* R577X (rs1815739) genetic variants in elite rugby athletes (rugby union and league) and compare genotype frequencies to controls and between playing positions. The rugby athlete cohort consisted of 507 Caucasian men, including 431 rugby union athletes that for some analyses were divided into backs and forwards and into specific positional groups: front five, back row, half backs, centers, and back three. Controls were 710 Caucasian men and women. Real-time PCR of genomic DNA was used to determine genotypes using TaqMan probes and groups were compared using χ^2 and odds ratio (OR) statistics. Correction of P values for multiple comparisons was according to Benjamini-Hochberg. There was no difference in *ACE* I/D genotype between groups. *ACTN3* XX genotype tended to be underrepresented in rugby union backs (15.7%) compared with forwards (24.8%, $P = 0.06$). Interestingly, the 69 back three players (wings and full backs) in rugby union included only six XX genotype individuals (8.7%), with the R allele more common in the back three (68.8%) than controls (58.0%; $\chi^2 = 6.672$, $P = 0.04$; OR = 1.60) and forwards (47.5%; $\chi^2 = 11.768$, $P = 0.01$; OR = 2.00). Association of *ACTN3* R577X with playing position in elite rugby union athletes suggests inherited fatigue resistance is more prevalent in forwards, while inherited sprint ability is more prevalent in backs, especially wings and full backs. These results also demonstrate the advantage of focusing genetic studies on a large cohort within a single sport, especially when intrasport positional differences exist, instead of combining several sports with varied demands and athlete characteristics.

α-actinin-3; angiotensin converting enzyme; athlete genetics; RugbyGene project

Address for reprint requests and other correspondence: S. M. Heffernan, MMU Sports Genomics Laboratory, Manchester Metropolitan Univ., Crewe Green Rd., Crewe, UK, CW1 5DU (e-mail: shane.heffernan@stu.mmu.ac.uk).

RUGBY IS AN INTERMITTENT TEAM sport comprising two similar but differing codes, rugby league (RL) and rugby union (RU). Both codes consist of diverse playing positions, each with different physiological, anthropometric, and technical attributes (8, 10, 20, 27) including two distinct subgroups in each code: forwards and backs. Recently, global positioning system tracking and time-motion analysis have been used to estimate the physical demands of rugby athletes and compare forwards and backs during high-level match play (8, 20, 27). In RU, backs travelled 12% greater total distance (6,545 m vs. 5,850 m), achieved maximum speeds 16% faster (30.4 km/h vs. 26.3 km/h) and engaged in over four times (58% vs. 13%) high-intensity running activities (>5.0 m/s), as a proportion of total activity (8, 27) compared with forwards. These data suggest a more sprint-oriented metabolic demand in backs compared with forwards. Furthermore, due to the complexities of forward play, forwards performed sixfold more (9.9%) high-intensity static exertion activities (rucks, mauls, scrums, and line-outs) than backs (1.6%) and spent 19.8% more time running above 80% of their maximal speed (8, 27, respectively). This implies that forwards, although often of higher body mass (14), are more likely to benefit from fatigue-resistant physiological qualities than backs. Accordingly, Deutsch et al. (10) showed that forwards had a notably higher work-to-rest ratio than backs (1:7 and 1:22, respectively). Given that the roles of backs and forwards differ significantly in terms of physiological demands, these differences may be reflected in distinct genetic characteristics (18). Elite RL athletes cover similar total distances (~7,000 m vs. ~5,000 m; backs vs. forwards, respectively) and have similar anthropometric characteristics to RU athletes (20). Players regularly transfer between RL and RU codes so investigating both codes (combined and separately) for their genetic characteristics is justified.

The two most studied gene variants in exercise genomics (*ACE* I/D and *ACTN3* R577X polymorphisms) have recently been considered in meta-analyses. Ma et al. (23) reported that *ACE* II genotype was associated with physical performance [odds ratio (OR) 1.23], especially endurance performance (OR

1.35). Furthermore, *ACTN3* RR genotype was associated with speed and power performance (OR 1.21; 23), supported elsewhere (2). More extensive information regarding *ACE* I/D and *ACTN3* R577X polymorphisms is available (13, 26). Due to differences in physical characteristics between rugby athletes and the general population and the diverse physiological demands within rugby, these genetic markers could predispose athletes to success or specific roles at the elite level.

One recent paper examined *ACE* I/D genotype frequency distribution in young, nonelite RU athletes. *ACE* I/D genotype frequencies did not differ between forwards and backs, with no control group included (5). The same group (4) also investigated *ACTN3* R577X in 102 young male RU athletes and reported no association, despite some tendencies for the R allele to be more frequent in backs or subgroups of backs. Studying elite athletes would be better able to answer the question whether these genetic variants are associated with elite status and playing position in rugby.

Therefore, the purpose of the present study was to investigate whether elite rugby athletes in the RugbyGene project (18) and a control group differed in terms of *ACE* I/D and *ACTN3* R577X genotype distribution and whether athletes in specialized playing positions similarly differed. It was hypothesized that the *ACTN3* R allele and the *ACE* I allele would be more frequent in rugby athletes than controls. It was further hypothesized that *ACTN3* XX and *ACE* II genotypes would be underrepresented in RU backs compared with forwards, due to differences in overall work-to-rest ratio and differing requirements for high maximum speed.

METHODS

Participants

Ethical approval was granted by Manchester Metropolitan University (MMU), University of Glasgow, University of Cape Town, and Northampton University ethics committees, and the study complies with the Declaration of Helsinki. As part of the RugbyGene project, elite Caucasian male rugby athletes [$n = 507$; mean (standard deviation) height 1.85 (0.07) m, mass 101 (14) kg, age 29 (7) yr] including 71.2% British, 17.2% South African, 7.1% Irish, and 4.5% of other nationalities were recruited, having given written informed consent. Caucasian controls [61% male; $n = 710$; height 1.73 (0.10) m, mass 74 (13) kg, age 29 (16) yr] included 89.6% British, 8.9% South African, 0.7% Irish, and 0.8% of other nationalities. Athletes were considered elite if they had competed regularly (>5 matches) since 1995 in the highest professional league in the UK, Ireland, or South Africa for RU and the highest professional league in the UK for RL. Of the RU athletes, 53.4% had competed at international level for a "High Performance Union" (Regulation 16, http://www.worldrugby.org), and 38.5% of RL had competed at international level. International status was confirmed as of 1 January 2015. Athletes were taller and heavier ($P < 0.0005$) but not older ($P = 0.871$) than controls.

Procedures

Sample collection. Blood (~70% of all samples), saliva (~25%), or buccal swab samples (~5%) were obtained via the following protocols. Blood was drawn from a superficial forearm vein into an EDTA tube and stored in sterile tubes at −20°C until processing. Saliva samples were collected into Oragene DNA OG-500 collection tubes (DNA Genotek, Ottawa, Ontario, Canada) according to the manufacturer's protocol and stored at room temperature until processing. Sterile buccal swabs (Omni swab; Whatman, Springfield Mill, UK) were rubbed against the buccal mucosa of the cheek for ~30 s. Tips were ejected into sterile tubes and stored at −20°C until processing.

DNA isolation and genotyping. DNA isolation and genotyping were performed in the MMU, University of Glasgow, University of Cape Town (DNA isolation only), and University of Northampton laboratories. There are some differences between protocols summarized below; however, there was 100% agreement among reference samples genotyped in the three genotyping centers, i.e., Glasgow, Northampton, and MMU laboratories. The majority of samples were processed and genotyped in the MMU laboratory. Genotype calling was successful for both variants in all samples.

At MMU and Glasgow, DNA isolation was performed with the QIAamp DNA Blood Mini kit and standard spin column protocol, following the manufacturer's instructions (Qiagen, West Sussex, UK). Briefly, 200 μl of whole blood/saliva, or one buccal swab, was lysed and incubated, the DNA washed, and the eluate containing isolated DNA stored at 4°C. In Cape Town, DNA was isolated from whole blood by a different protocol (22). In brief, samples were lysed and centrifuged, the DNA washed, and samples stored at −20°C. Genotyping of DNA isolated in Cape Town was performed in Glasgow. At Northampton, DNA was isolated from whole blood with Flexigene kits (Qiagen). In brief, samples were lysed, and DNA precipitated and washed, with samples stored at −20°C.

Genotyping. Genotyping in the Glasgow laboratory was performed on *ACTN3* (rs1815739) and an *ACE* tag single nucleotide polymorphism (SNP) (rs4341) in perfect linkage disequilibrium with *ACE* I/D in Caucasians (15). In brief, 10 μl Genotyping Master Mix (Applied Biosystems, Paisley, UK), 1 μl SNP-specific TaqMan assay (Applied Biosystems), 6 μl nuclease-free H_2O, and 3 μl DNA solution (~9 ng DNA) were added per well. In the Northampton laboratory, genotyping was performed for *ACTN3* R577X (rs1815739) by combining 10 μl of Genotyping Master Mix, 8 μl H_2O, 1 μl assay mix with 1 μl of purified DNA (~10 ng). In both laboratories, PCR was performed using a StepOnePlus real-time detector (Applied Biosystems). In brief, denaturation began at 95°C for 10 min, with 40 cycles of incubation at 92°C for 15 s then annealing and extension at 60°C for 1 min. Initial analysis was performed using StepOnePlus software version 2.3 (Applied Biosystems). There was 100% agreement within duplicates of all samples.

At MMU, samples were genotyped for *ACTN3* R577X (rs1815739) by combining 5 μl Genotyping Master Mix, 4.3 μl H_2O, 0.5 μl assay mix, and 0.2 μl of purified DNA (~9 ng), for samples derived from blood and saliva. For DNA derived from buccal swabs, 5 μl Genotyping Master Mix was combined with 3.5 μl H_2O, 0.5 μl assay mix, and 1 μl DNA solution (~9 ng DNA). Either a Chromo4 real-time system (Bio-Rad, Hertfordshire, UK) or a StepOnePlus was used. In brief, denaturation began at 95°C for 10 min, with 40 cycles of incubation at 92°C for 15 s and then annealing and extension at 60°C for 1 min. Initial genotyping analysis was performed with Opticon Monitor software version 3.1 (Bio-Rad) or StepOnePlus software version 2.3. Duplicates of all samples were in 100% agreement. For *ACE* I/D at MMU, 5 μl of Genotyping Master Mix, 1.55 μl H_2O, 0.9 μl of I and D allele-specific probes, and 0.38 μl of *ACE* primer 111, 112, and 113 (sequences below) were combined with 0.5 μl DNA solution (~23 ng DNA) per well for blood and saliva. For DNA derived from buccal cells, primer and probe volumes were identical, but 0.05 μl H_2O and 2 μl DNA solution (~18 ng DNA) were used. Similarly, in the Northampton laboratory, *ACE* I/D was genotyped by combining 11 μl of Genotyping Master Mix, 2 μl of I and D probes, 2 μl of *ACE* primer 111, 112, 113, and 4 μl DNA solution (~40 ng DNA). Either a Chromo4 real-time system or a StepOnePlus was used. In brief, there were 50 cycles of denaturation at 92°C for 15 s and then annealing and extension at 57°C for 1 min. Initial analysis was performed with Opticon Monitor 3.1 software or StepOnePlus software version 2.3. Again, there was 100% agreement within duplicates of all samples.

Primers and probes. For rs1815739 and rs4341, the appropriate TaqMan assay was used (Applied Biosystems). For the direct *ACE* I/D assay, three primers (150 nM each) and probes (VIC, 150 nM and FAM, 75 nM; 21) were used: primer ACE111, 5′-CCCATCCTTTCTC-CCATTTCTC-3′; primer ACE112, 5′-AGCTGGAATAAAATTGGC-GAAAC-3′; primer ACE113, 5′-CCTCCCAAAGTGCTGGGATTA-3′; I allele-specific probe (VIC-ACE100), VIC-5′-AGGCGTGATA-CAGTCA-3′-MGB; D allele-specific probe (FAM-ACE100), FAM-5′-TGCTGCCTATACAGTCA-3′-MGB.

Positional Groups

To assess genotype and allele frequencies within the RU group, we allocated athletes to subgroups: forwards (props, hookers, locks, flankers, number eights) and backs (scrum halves, fly halves, centers, wings, full backs). Also, due to diverse physiological demands within RU (8, 27), athletes were further divided into positional groups according to their similar movement patterns (8) front five (props, hookers, locks), back row (flankers, number eights), half backs (scrum halves, fly halves), centers, and back three (wings and full backs). Comparisons between positions were not performed for the RL cohort due to low statistical power when it was subdivided.

Data Analysis

SPSS for Windows version 19 (SPSS, Chicago, IL) software was used to conduct Pearson's Chi-square (χ^2) tests to compare genotype and allelic frequencies between athletes and controls and between positional subgroups. For *ACTN3* and *ACE*, 26 and 16 tests, respectively, were subjected to Benjamini-Hochberg (BH; 6) corrections to control false discovery rate, and corrected probability values are reported. Where appropriate, OR was calculated to estimate effect size. Alpha was set at 0.05.

RESULTS

All genotype data for athletes and controls were in Hardy-Weinberg equilibrium. There were no differences in genotype frequencies within the athlete or control groups according to nationality. For *ACE* I/D, there were no differences between all athletes (RU and RL combined) and controls in genotype (χ^2 = 1.117, P = 0.83), between RU or RL and controls, nor between playing subgroups for RU (Table 1). Furthermore, for *ACTN3* R577X there were no genotype differences between controls and all athletes (χ^2 = 1.645, P = 0.44), RL (χ^2 = 1.829, P = 0.44), or RU (χ^2 = 0.216, P = 0.33). However, when we considered RU playing position, the X allele was overrepresented in forwards (52.5%) compared with backs [37.8%, χ^2 = 8.128, P = 0.02; OR = 1.49, 95% confidence interval (CI) = 1.13–1.96, P = 0.004] and controls (42%, χ^2 = 6.217, P = 0.02; OR = 1.25, 95% CI = 1.02–1.54, P = 0.033; Table 1 and Fig. 1*A*). Similarly, there was a tendency (P = 0.023 before BH correction) of the XX genotype to be overrepresented in forwards (24.8%) compared with backs (15.7%, χ^2 = 5.193, P = 0.08; OR = 1.77, 95% CI = 1.09–2.89, P = 0.022) and controls (18.3%, χ^2 = 7.582, P = 0.08), with no difference between backs and controls (χ^2 = 3.043, P = 0.37).

Interestingly, the 69 back three athletes (wings and full-backs) included only six individuals (8.7%) of XX genotype that differed from the forwards (24.8%; χ^2 = 11.082, P = 0.05; OR = 3.46, 95% CI = 1.43–8.34, P = 0.006) and tended to differ from the combined half backs and centers group (19.8%; χ^2 = 4.151, P = 0.08; OR = 2.59, 95% CI = 1.00–6.74, P = 0.049). Likewise, the R allele distribution was greater in the back three (68.8%) than the controls (58.0%; χ^2 = 6.672, P = 0.02; OR = 1.60, 95% CI = 1.09–2.33, P = 0.014), forwards (47.5%; χ^2 = 11.768, P = 0.01; OR = 2.00, 95% CI = 1.34–2.99, P = 0.0007), and the other backs (58.2%; χ^2 = 4.173, P = 0.05; OR = 1.59, 95% CI = 1.02–2.48, P = 0.042) (Fig. 1*B*).

DISCUSSION

The present study is the first to show a genetic association with elite athlete status in RU. We found associations for the *ACTN3* R577X polymorphism but not for *ACE* I/D, thus rejecting our hypotheses regarding *ACE* I/D. Furthermore, no difference was observed for the *ACTN3* R577X genotype or allele distribution between all athletes and controls, thus rejecting the hypothesis that differences would exist between nonathletes and all players as a single cohort. Similarly, there were no differences between the RU, RL, and control groups when playing position was not considered. However, as hypothesized, in RU backs compared with forwards there was a lower proportion of XX genotype and X allele, which probably reflects the greater need for speed generation in backs and more sustained activity in forwards. The small cohort of RL athletes

Table 1. *Genotype and allele distribution of controls and athletes divided into positional subgroups (for RU only), presented as genotype/allele counts followed by percentage in parentheses*

Genotype	All Athletes	RL Athletes	RU Athletes	Controls	Forwards	Front 5	Back Row	Backs	Half Backs	Centers	Back 3
					ACE						
II	108 (21.4)	18 (21.7)	92 (21.5)	113 (19.8)	49 (20.0)	36 (22.1)	13 (15.9)	43 (23.6)	14 (20.3)	14 (31.1)	15 (22.1)
ID	251 (49.7)	39 (47.0)	214 (50.1)	286 (50.0)	129 (52.7)	86 (52.8)	43 (52.4)	85 (46.7)	33 (47.8)	17 (37.8)	35 (51.5)
DD	146 (28.9)	26 (31.3)	121 (28.3)	172 (30.2)	67 (27.3)	41 (25.2)	26 (31.7)	54 (29.7)	22 (31.9)	14 (31.1)	18 (26.5)
Total	505	83	427	572	245	163	82	182	69	45	68
I allele	467 (46.3)	75 (45.2)	398 (46.6)	512 (44.7)	227 (46.3)	158 (48.5)	69 (42.1)	171 (47.0)	61 (44.2)	45 (50.0)	65 (47.8)
D allele	543 (53.7)	91 (54.8)	456 (53.4)	630 (55.3)	263 (53.7)	168 (51.5)	95 (57.9)	193 (53.0)	77 (55.8)	45 (50.0)	71 (52.2)
					ACTN3						
XX	104 (20.5)	15 (18.1)	90 (20.9)	130 (18.3)	61 (24.8)	39 (23.8)	22 (26.8)	29 (15.7)	12 (17.4)	11 (23.4)	6 (8.7)*
RX	234 (46.2)	45 (54.2)	194 (45.0)	337 (47.5)	112 (45.5)	71 (43.3)	41 (50.0)	82 (44.3)	29 (42.0)	22 (46.8)	31 (44.9)
RR	169 (33.3)	23 (27.7)	147 (34.1)	243 (34.2)#	73 (29.7)#	54 (32.9)	19 (23.2)	74 (40.0)	28 (40.6)	14 (29.8)	32 (46.4)
Total	507	83	431	710	246	164	82	185	69	47	69
X allele	442 (43.5)	75 (45.2)	374 (43.4)	597 (42.0)*	234 (47.6)	149 (45.4)	85 (51.8)	140 (37.8)*	53 (38.4)	44 (46.8)	43 (31.2)
R allele	572 (56.5)	91 (54.8)	488 (56.6)	823 (58.0)#	258 (52.4)	179 (54.6)	79 (48.2)	230 (62.2)	85 (61.6)	50 (53.2)	95 (68.8)*

RL, rugby league; RU, rugby union. *Different from forwards. #Different from the Back 3.

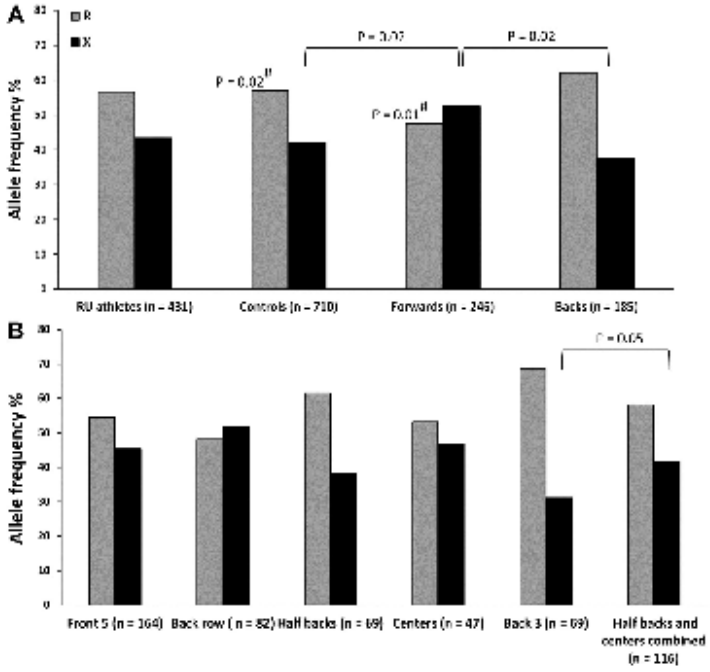


Fig. 1. *ACTN3* allele frequencies. *A*: allele frequencies of rugby union (RU) athletes and controls, with athletes also divided into playing subgroup (forwards and backs). #Different from the back 3. *B*: allele frequencies of RU athletes divided into positional groups with the addition of the "half backs and centers combined" group. Statistical analysis between these positional groups only compared the back 3 with the half backs and centers combined.

means that comparisons between playing positions are not feasible until the cohort increases substantially.

ACTN3 R577X

The most remarkable finding of the present study was the low frequency of the XX genotype among the back three RU athletes (8.7%), approaching, although not as low as, the frequency observed in elite sprinters (25, 31). The XX genotype is present in ~18% of Caucasians (Table 1) and indicates absence of the α-actinin-3 protein (3, 24). Absence of α-actinin-3, a protein almost exclusively expressed in fast-twitch skeletal muscle fibers, could hinder back three (wing and full back) sprint ability. R allele carriers have a greater proportion of type II and IIx fibers and larger relative surface area per IIx fiber than XX carriers (1, 7, 30). Furthermore, Seto et al. (29) recently showed the likely mechanism for this genotype-phenotype association is via the calcineurin muscle fiber remodeling pathway. They found greater calcineurin activity (which induces slow myogenic programming and a shift toward oxidative phenotype) in α-actinin-3 knockout mice (KO) and humans (*ACTN3* 577XX genotype) due to preferential binding of α-actinin-2 (upregulated in the absence of α-actinin-3) to the fast fiber-specific calsarcin-2 (an inhibitor of calcineurin). This could explain the advantage of R allele carriers over α-actinin-3-deficient XX individuals for high-velocity contractions, particularly important for back three RU players. While backs and forwards previously showed similar fiber type proportions (19), these older data are arguably not relevant to modern rugby athletes, given their changed physical characteristics in recent years (14). Skeletal muscle fiber type proportions are unknown in contemporary elite RU athletes who now compete in a more popular, fully professional sport and complete much higher training loads than previously. Recent in vivo data also show that R allele carriers exhibit greater muscle volume and maximal power output (11, 17). While RU forwards show greater maximal power, backs are able to generate greater power relative to body mass (W/kg; 9), which corresponds with the greater R allele frequency in the backs and especially the back three players. These data, plus evidence that type II fibers are larger and more powerful per unit volume than type I (15), suggest the R allele would benefit back three rugby athletes for muscle power and fast fiber characteristics, which supports our findings (Table 1 and Fig. 1).

Arguably, the higher propensity for aerobic enzyme activity (porin, COX IV, hexokinase, citrate synthase, succinate dehydrogenase, and β-hydroxyacyl CoA dehydrogenase; 28, 29) and greater force recovery after fatigue observed in α-actinin-3-deficient mice (28) could indicate that XX genotype humans might have a greater capacity for recovery from fatiguing exercise, a trait that would benefit forwards with their more sustained match play intensity and necessity for quick recovery. The shorter rest periods for forwards compared with backs (work-to-rest ratios 1:7.4 and 1:21.8, respectively; 10) indicates that greater fatigue resistance would be particularly beneficial for forwards. Moreover, the greater calcineurin activity in XX homozygote humans and approximately threefold

increase in calcineurin activity and distance run after endurance training in KO mice (29) further support the notion that forwards would benefit from a greater fatigue resistance, especially with exposure to extensive training. These data are consistent with our observation that forwards exhibit higher XX genotype and lower R allele frequencies than backs and controls (Table 1).

When many sports were considered simultaneously, team sport athlete status showed no association with *ACTN3* R577X genotype (12). However, when one considers the relatively small number of athletes (205) with mixed status (56.6% elite) from a range of sports (ice hockey, handball, soccer, etc.), that is perhaps not surprising. While combining cohorts from different sports can boost sample size and theoretically increase statistical power, if an association does not exist in all sports, or even in all athletes within a particular sport due to positional differences, one would be less likely to detect an association. The positional differences identified within the present study demonstrate the value of studying a large sample from a single sport and, in the absence of detailed physiological data (often difficult to obtain from large numbers of elite athletes), provide a viable alternative for future genetic research involving team sport athletes.

ACE I/D

The current study reports no difference between rugby athletes and controls or any positional subgroups for *ACE* I/D. This lack of association contrasts with a recent meta-analysis where the *ACE* I allele was associated with physical performance (23). However, Ma et al. (23) also report that specialized distance/endurance athletes showed the strongest association with the I allele (OR 1.35). Given the mixed metabolic nature of rugby, a comparable association in the present study was less likely. Furthermore, the importance of *ACE* I/D remains controversial in the literature, with no associations reported in other isolated team sports such as elite European soccer (16) and nonelite RU (5). These prior data, in conjunction with our current findings in a larger study that also considers playing position, suggest that *ACE* I/D plays little role in performance of team sport athletes. *ACE* I/D genotype-athlete phenotype associations are more likely to exist in specialized endurance athletes (26).

Effect Size and Future Applications

OR were calculated to estimate the likelihood that individuals with the advantageous genotype/allele become an elite RU athlete in a specific position. The *ACTN3* XX genotype was almost twice (OR = 1.77) as common in forwards than backs, which suggests α-actinin-3-deficient individuals are more suited to forward play. Furthermore, forwards were over three times (OR = 3.46) more likely to be XX genotype than the back three athletes, while the remaining backs (centers and halves) were over twice as likely to show the α-actinin-3-deficient genotype than the back three (OR = 2.59). These data suggest the *ACTN3* R577X polymorphism shows potential to contribute to position-specific player profiling within RU when combined with other genetic and physiological data in the future. In contrast, the *ACE* I/D polymorphism (OR ~1) does not show equivalent potential in rugby.

While the present cohort size is large compared with previous single sport genetic analyses, when the cohort was subdivided into playing position, the numbers were reduced so enlargement of our cohort and replication would be welcome. Accordingly we continue to recruit elite RU and RL players in the RugbyGene project and so will steadily become better able to investigate genetic aspects of specific demands within rugby. To conclude, the present study revealed position-specific genetic variation in elite RU athletes for *ACTN3* R577X. The R allele was an advantage for backs, particularly the back three. Moreover, the current results do not support *ACE* I/D as a genetic marker for rugby performance, showing no differences between athletes and controls or positional subgroups. This study demonstrates the value of single sport cohorts and the need for large sample sizes when conducting gene association studies in sport. Future objectives of the RugbyGene project within the broader Athlome project include investigating whether genetic variants associated with excellence in other sports are similarly associated in the multifaceted sport of rugby.

ACKNOWLEDGMENTS

The authors thank all athletes, their respective scientific support staff, and the control participants for their time and willingness to participate in this research. We thank Louis El Khoury for technical assistance at the University of Northampton and Ben Davies for assistance in participant recruitment.

Twitter: Follow the RugbyGene project at @RugbyGeneStudy.

DISCLOSURES

No conflicts of interest, financial or otherwise, are declared by the author(s).

AUTHOR CONTRIBUTIONS

Author contributions: S.M.H., L.P.K., R.M.E., S.H.D., M.C., Y.P.P., and A.G.W. conception and design of research; S.M.H., R.M.E., J.S.M., G.K.S., J.P.H.N., S.J.I., M.C., and A.G.W. performed experiments; S.M.H., R.M.E., J.S.M., G.K.S., J.P.H.N., S.J.I., M.C., and A.G.W. analyzed data; S.M.H., L.P.K., R.M.E., S.H.D., J.S.M., G.E.M., G.K.S., J.P.H.N., S.J.I., W.J.R., C.J.C., B.V., S.M.R., C.R., M.A.B., G.W., M.C., Y.P.P., and A.G.W. interpreted results of experiments; S.M.H. prepared figures; S.M.H. and C.R. drafted manuscript; S.M.H., L.P.K., R.M.E., S.H.D., J.S.M., G.E.M., G.K.S., J.P.H.N., S.J.I., W.J.R., C.J.C., B.V., S.M.R., C.R., M.A.B., G.W., M.C., Y.P.P., and A.G.W. edited and revised manuscript; S.M.H., L.P.K., R.M.E., S.H.D., J.S.M., G.E.M., G.K.S., J.P.H.N., S.J.I., W.J.R., C.J.C., B.V., S.M.R., C.R., M.A.B., G.W., M.C., Y.P.P., and A.G.W. approved final version of manuscript.

REFERENCES

1. **Ahmetov II, Druzhevskaya AM, Lyubaeva EV, Popov DV, Vinogradova OL, Williams AG.** The dependence of preferred competitive racing distance on muscle fibre type composition and ACTN3 genotype in speed skaters. *Exp Physiol* 96: 1302–1310, 2011.
2. **Alfred T, Ben-Shlomo Y, Cooper R, Hardy R, Cooper C, Deary IJ, Gunnell D, Harris SE, Kumari M, Martin RM, Moran CN, Pitsiladis YP, Ring SM, Sayer AA, Smith GD, Starr JM, Kuh D, Day IN.** ACTN3 genotype, athletic status, and life course physical capability: meta-analysis of the published literature and findings from nine studies. *Hum Mutat* 32: 1008–1018, 2011.
3. **Beggs AH, Byers TJ, Knoll JH, Boyce FM, Bruns GA, Kunkel LM.** Cloning and characterization of two human skeletal muscle alpha-actinin genes located on chromosomes 1 and 11. *J Biol Chem* 267: 9281–9288, 1992.
4. **Bell W, Colley JP, Evans WD, Darlington SE, Cooper SM.** ACTN3 genotypes of rugby union players: distribution, power output and body composition. *Ann Hum Biol* 39: 19–27, 2012.
5. **Bell W, Colley JP, Gwynne JR, Glazier P, Evans WD, Darlington SE.** ACE ID genotype and leg power in rugby union players. *J Sports Med Phys Fit* 50: 350–355, 2010.

6. **Benjamini Y, Hochberg Y.** Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J Royal Stat Soc* 57: 289–300, 1995.
7. **Broos S, Malisoux L, Theisen D, Francaux M, Deldicque L, Thomis MA.** Role of alpha-actinin-3 in contractile properties of human single muscle fibers: a case series study in paraplegics. *PLoS One* 7: 8, 2012.
8. **Cahill N, Lamb K, Worsfold P, Headey R, Murray S.** The movement characteristics of English premiership rugby union players. *J Sports Sci* 31: 229–237, 2012.
9. **Crewther B, Kilduff L, Cook C, Cunningham D, Bunce P, Bracken R, Gaviglio C.** Scaling strength and power for body mass differences in rugby union players. *J Sport Med Phys Fit* 52: 27–32, 2012.
10. **Deutsch MU, Kearney GA, Rehrer NJ.** Time-motion analysis of professional rugby union players during match-play. *J Sports Sci* 25: 461–472, 2007.
11. **Erskine RM, Williams AG, Jones DA, Stewart CE, Degens H.** The individual and combined influence of ACE and ACTN3 genotypes on muscle phenotypes before and after strength training. *Scand J Med Sci Sport* 24: 642–664, 2013.
12. **Eynon N, Banting LK, Ruiz JR, Cieszczyk P, Dyatlov DA, Maciejewska-Karlowska A, Sawczuk M, Pushkarev VP, Kulikov LM, Pushkarev ED, Femia P, Stepto NK, Bishop DJ, Lucia A.** ACTN3 R577X polymorphism and team-sport performance: A study involving three European cohorts. *J Sci Med Sport* 17: 102–106, 2013.
13. **Eynon N, Hanson ED, Lucia A, Houweling PJ, Garton F, North KN, Bishop DJ.** Genes for elite power and sprint performance: ACTN3 leads the way. *Sports Med* 43: 803–817, 2013.
14. **Fuller CW, Taylor AE, Brooks JH, Kemp SP.** Changes in the stature, body mass and age of English professional rugby players: a 10-year review. *J Sports Sci* 31: 795–802, 2013.
15. **Gilliver SF, Degens H, Rittweger J, Sargeant AJ, Jones DA.** Variation in the determinants of power of chemically skinned human muscle fibres. *Exp Physiol* 94: 1070–1078, 2009.
16. **Gineviciene V, Jakaitiene A, Tubelis L, Kucinskas V.** Variation in the ACE, PPARGC1A and PPARA genes in Lithuanian football players. *Eur J Sport Sci* 14: S289–S295, 2014.
17. **Gómez-Gallego F, Santiago C, González-Freire M, Muniesa CA, Del Valle MF, Perez M, Foster C, Lucia A.** Endurance performance: genes or gene combinations? *Int J Sports Med* 30: 66–72, 2009.
18. **Heffernan SM, Kilduff LP, Day SH, Pitsiladis YP, Williams AG.** Genomics in rugby union: a review and future prospects. *Eur J Sport Sci* 1–9, 2015.
19. **Jardine MA, Wiggins TM, Myburgh KH, Noakes TD.** Physiological characteristics of rugby players including muscle glycogen content and muscle fibre composition. *S Afr Med J* 73: 529–532, 1988.
20. **Johnston RD, Gabbett TJ, Jenkins DG.** Applied sport science of rugby league. *Sports Med* 44: 1087–1100, 2014.
21. **Koch W, Latz W, Eichinger M, Ganser C, Schömig A, Kastrati A.** Genotyping of the angiotensin I-converting enzyme gene insertion/deletion polymorphism by the TaqMan method. *Clin Chem* 51: 1547–1549, 2005.
22. **Lahiri DK, Nurnberger JI.** A rapid non-enzymatic method for the preparation of HMW DNA from blood for RFLP studies. *Nucleic Acids Res* 19: 5444–5445, 1991.
23. **Ma F, Yang Y, Li X, Zhou F, Gao C, Li M, Gao L.** The association of sport performance with ACE and ACTN3 genetic polymorphisms: A systematic review and meta-analysis. *PLoS One* 8: e54685, 2013.
24. **Mills M, Yang N, Weinberger R, Vander Woude DL, Beggs AH, Easteal S, North K.** Differential expression of the actin-binding proteins, α-actinin-2 and-3, in different species: implications for the evolution of functional redundancy. *Hum Mol Genet* 10: 1335–1346, 2001.
25. **Niemi AK, Majamaa K.** Mitochondrial DNA and ACTN3 genotypes in Finnish elite endurance and sprint athletes. *Eur J Hum Genet* 13: 965–969, 2005.
26. **Puthucheary Z, Skipworth JRA, Rawal J, Loosemore M, Van Someren K, Montgomery HE.** The ACE gene and human performance. *Sport Med* 41: 433–448, 2011.
27. **Roberts SP, Trewartha G, Higgitt RJ, El-Abd J, Stokes KA.** The physical demands of elite English rugby union. *J Sport Sci* 26: 825–833, 2008.
28. **Seto JT, Chan S, Turner N, MacArthur DG, Raftery JM, Berman YD, Quinlan KG, Cooney GJ, Head S, Yang N, North KN.** The effect of alpha-actinin-3 deficiency on muscle aging. *Exp Gerontol* 46: 292–302, 2011.
29. **Seto JT, Quinlan KGR, Lek M, Zheng XF, Garton F, MacArthur DG, Hogarth MW, Houweling PJ, Gregorevic P, Turner N, Cooney GJ, Yang N, North KN.** ACTN3 genotype influences muscle performance through the regulation of calcineurin signaling. *J Clin Invest* 123: 4255–4263, 2013.
30. **Vincent B, De Bock K, Ramaekers M, Van den Eede E, Van Leemputte M, Hespel P, Thomis MA.** ACTN3 (R577X) genotype is associated with fiber type distribution. *Physiol Genomics* 32: 58–63, 2007.
31. **Yang N, MacArthur DG, Gulbin JP, Hahn AG, Beggs AH, Easteal S, North K.** ACTN3 genotype is associated with human elite athletic performance. *Am J Hum Genet* 73: 627–631, 2003.

European Journal of Sport Science, 2015
Vol. 15, No. 6, 460–468, http://dx.doi.org/10.1080/17461391.2015.1023222

REVIEW ARTICLE

Genomics in rugby union: A review and future prospects

SHANE M. HEFFERNAN[1], LIAM P. KILDUFF[2], STEPHEN H. DAY[1], YANNIS P. PITSILADIS[3], & ALUN G. WILLIAMS[1,4]

[1]*MMU Sports Genomics Laboratory, Manchester Metropolitan University, Crewe, UK,* [2]*A-STEM, College of Engineering, Swansea University, Swansea, UK,* [3]*Centre for Sport and Exercise Science and Medicine (SESAME), University of Brighton, Brighton, UK,* [4]*Institute of Sport, Exercise and Health, University College London, London, UK*

Abstract
This article introduces some aspects of sports genomics in a rugby union context, considers the rugby-specific genetic data in the published literature and outlines the next research steps required if the potential applications of genetic technology in rugby union, also identified here, are to become possible. A substantial proportion of the inter-individual variation for many traits related to rugby performance, including strength, short-term muscle power, $VO_{2\ max}$, injury susceptibility and the likelihood of being an elite athlete is inherited and can be investigated using molecular genetic techniques. In sports genomics, significant efforts have been made in recent years to develop large DNA biobanks of elite athletes for detailed exploration of the heritable bases of those traits. However, little effort has been devoted to the study of rugby athletes, and most of the little research that has focused on rugby was conducted with small cohorts of non-elite players. With steadily growing knowledge of the molecular mechanisms underpinning complex performance traits and the aetiology of injury, investigating sports genomics in the context of rugby is now a viable proposition and a worthwhile endeavour. The RugbyGene project we describe briefly in this article is a multi-institutional research collaboration in rugby union that will perform molecular genetic analyses of varying complexity. Genetic tests could become useful tools for rugby practitioners in the future and provide complementary and additional information to that provided by the non-genetic tests currently used.

Keywords: *Sports genetics, athlete status, rugby physiology*

Introduction

The majority of scientific investigation into player performance in rugby union has focused on environmental factors such as training methods, dietary supplementation and recovery strategies (Barr, Sheppard, Gabbett, & Newton, 2014; Bradley et al., 2014; West et al., 2013), with a great deal of epidemiological research also investigating injury frequency and risk including a recent focus on brain injury (Gardner, Iverson, Williams, Baker, & Stanwell, 2014; Raftery, 2014). However, considerable evidence shows that performance and injury traits are highly (but variably) heritable (Bouchard et al., 1999; De Moor et al., 2007; Hakim, Cherkas, Spector, & MacGregor, 2003; Peeters et al., 2007; Simoneau & Bouchard, 1995), yet little scientific effort has been invested to elucidate this inter-individual variation within rugby union. Rugby union athletes are qualitatively and quantitatively different from other athlete groups (both individual and team), in that there are vast difference in the physiological and anthropometric characteristics across a single rugby union team according to playing position (Smart, Hopkins, & Gill, 2013). Rugby union is also distinctive in that individual clusters of positions require different movement patterns in elite players (Quarrie, Hopkins, Anthony, & Gill, 2012) and thus differ in their metabolic demands.

For example, anthropometric and physiological variables differ significantly across playing positions, with elite scum half players averaging ~177 cm and ~85 kg, in contrast to props averaging ~185 cm and ~117 kg – a difference of 8 cm and 32 kg (Fuller, Taylor, Brooks, & Kemp, 2013). Furthermore, in terms of positional specific physiological differences that may be reflected in players' genetic variation, backs show lower maximal strength compared to forwards in terms of bench press (difference ~11 kg),

Correspondence: Shane M. Heffernan, MMU Sports Genomics Laboratory, Manchester Metropolitan University, Crewe Green Road, Crewe CW1 5DU, UK. E-mail: shane.heffernan@stu.mmu.ac.uk

back squat (difference ~18 kg) and power clean (difference ~9 kg; Smart, Hopkins, Quarrie, & Gill, 2011). However, backs are faster sprinting 10 m (difference ~0.09 s) and 20 m (difference ~0.11 s) than forwards (Smart et al., 2011) and these differences become larger when specific positions are considered (Smart et al., 2013). Positional differences are further evidenced by game demand data that show the requirement for differing metabolic capacities dependent on playing position. Backs travel ~12% greater total distance (backs 6545 m, forwards 5850 m), achieving maximum speeds ~15% greater (forwards 26.3 km h^{-1}, backs 30.4 km h^{-1}) and are engaged in two to three times more high-intensity running than forwards (Cahill, Lamb, Worsfold, Headey, & Murray, 2012; Roberts, Trewartha, Higgitt, El-Abd, & Stokes, 2008). Furthermore, forwards perform more (~10%) high-intensity static exertion activities (rucks, mauls, scrums and line-outs) than backs (1.6%) and spend ~20% more time running faster than 80% of their maximal speed (Cahill et al., 2012; Roberts et al., 2008, respectively). The back three players (wings and fullbacks) travel the greatest distance not only at the lowest movement speeds (0–2 m s^{-1}) but also at the highest speeds (>8 m s^{-1}). In contrast, front row forwards (props and hookers) travel the greatest distance at moderately low (but not the lowest) movement speeds (2–4 m s^{-1}) and the least distance at high speeds (>8 m s^{-1}; Quarrie et al., 2012). These movement patterns correspond with the metabolic demands necessary to perform at the highest level in a given playing position within rugby union. The back three movement pattern suggests a requirement for a high proportion of fast twitch muscle fibres and highly developed anaerobic energy metabolism, with the front row movement pattern suggesting a greater capacity for rapid recovery between high-intensity static exertion activities. Given the highly heritable nature of these and other physical phenotypes that contribute to success in a given playing position, the differences should be reflected in distinct genetic characteristics. Furthermore, since rugby union athletes perform under a well-defined set of rules and parameters, which are ubiquitous across all playing positions, they present an ideal cohort via which to study the importance of genetic variation in sport.

In the present review, we will introduce the field of sports genomics, briefly review the current genetic research available for consideration within rugby union and outline the steps necessary to progress genomic understanding and applications in rugby union. Specifically, we will discuss some candidate genes that have been associated with physiological and anthropometric characteristics, injury risk, skill acquisition and athlete status. We will also introduce the highly collaborative RugbyGene project.

Human genomic variation

The variation that exists in the human genome has only relatively recently be documented via progress of the Human Genome Project. Initially, ~11 million single nucleotide polymorphisms (SNPs) and 3 million short insertions and deletions were identified (Frazer et al., 2007; Sachidanandam et al., 2001; The International HapMap Consortium, 2005) and, at the time of writing, more than 88 million SNPs have been validated (http://www.ncbi.nlm.nih.gov/projects/SNP/snp_summary.cgi). The vast majority of the ~3 billion nucleotides that constitute a human genome do not code for proteins, although most of those nucleotides nonetheless contribute to biological function in some manner via regulation of gene expression (ENCODE Project Consortium, 2012). There are a number of different types of human genomic variation that affect biological function, including rare mutations, structural variations and common SNPs.

The loss of function nucleotide change in the myostatin (*MSTN*) gene is an example of a rare mutation of relevance to human physiological function and results in an absence of the myostatin protein, which is a negative regulator of muscle growth. In humans, only one report of this mutation exists, causing extreme hyper-muscularity and a lean phenotype. At six days old, the affected child showed twice the quadriceps muscularity (cross-sectional area) than that of an age-matched control (6.7 cm^2 compared to 3.1 cm^2; Schuelke et al., 2004). The negative impacts (if any) of this mutation have not yet been reported and at 4.5 years the child showed no pathological symptoms. However, because myostatin affects other muscle types including cardiac muscle, later-onset pathological symptoms resulting from excessive myocardial growth are a possibility.

Genetic variations where an allele occurs relatively commonly (≥1%) are called polymorphisms, as opposed to mutations. An example of one such polymorphism relevant to exercise physiology is the insertion/deletion (I/D) variation on intron 16 of the *ACE* gene on chromosome 17. The *ACE* gene encodes the angiotensin converting enzyme, which is the main active product of the rennin–angiotensin system. The insertion of the 287 base pair (bp) DNA sequence (I allele), despite being in a non-protein-coding region of the *ACE* gene, produces lower circulating and tissue concentrations of the ACE protein compared to the D allele (Almeida et al., 2010). One main action of the ACE enzyme is to degrade inactive angiotensin I and generate the vasoconstrictor angiotensin II, while another is to degrade vasodilator kinins and thus is hypothesised to influence mitochondrial oxygen consumption

and exercise economy in some circumstances (Puthucheary et al., 2011).

An example of a SNP relevant to exercise performance found in an exon (a region of DNA that encodes for protein) is the *ACTN3* R577X gene variant. The *ACTN3* gene encodes for the α-actinin-3 protein, which is expressed almost exclusively in fast glycolytic type II fibres and is a structural component that binds the actin thin filament to the Z line. This SNP is located on exon 15 of the *ACTN3* gene (North et al., 1999) and is characterised by the replacement of the normal codon [a 3 bp sequence that transcribes for an amino acid (Arg; R)] by a premature termination codon (X) at the 577 amino acid position and results in the complete absence of the α-actinin-3 protein. The absence of this protein (XX genotype) is associated with a lower proportion of Type II muscle fibres and, accordingly, is found at a lower frequency in elite power/sprint athletes compared to other athletes and non-athletes (Eynon et al., 2013).

Physiological and anthropometric characteristics

Classical genetics is the process of estimating the heritability of a given trait and is investigated by studying families, identical twins (monozygotic – MZ) and fraternal twins (dizygotic – DZ). A number of physiological variables have been investigated in this way and are usually the preceding step to investigating specific target genes for further analysis. For example, Simoneau and Bouchard (1995) showed the genetic heritability of muscle fibre type proportion was ~50% when investigating 58 DZ and 35 MZ twin pairs. Furthermore, following 20 weeks of endurance training in 98 two-generation families (total $n = 481$), the heritable component of the ability to adapt, specifically an increase in $VO_{2\ max}$, was ~50% in the well-known HERITAGE family study (Bouchard et al., 1999). While these heritability estimates are substantial, other anthropometric phenotypes are estimated to be even greater: for example, body mass index at ~60%, height at ~80% (Silventoinen, Magnusson, Tynelius, Kaprio, & Rasmussen, 2008) and mesomorphy ~80% (Peeters et al., 2007).

Subsequently, molecular genetic technologies have been applied to examine the specific genes and their variations that might explain the heritable component of these and other phenotypes. For example, two intronic SNPs in the *TRHR* gene (rs7832552 and rs16892496) collectively account for ~5 kg of the variability in lean body mass (Liu et al., 2009). The *TRHR* gene encodes for the thyrotrophin-releasing hormone receptor that when bound to thyrotrophin-releasing hormone, results in a signalling cascade which leads to the release of thyroxin, an important component in the development of mammalian skeletal muscle. This gene variant, especially in combination with many others, could help explain a notable proportion of the ~80% heritability of mesomorphy mentioned previously and could be applied to the assessment of rugby players. Similarly, with regard to $VO_{2\ max}$, Bouchard et al. (2011) found the *PRDM1* gene accounted for ~7% of the inter-individual variability in adaptation to endurance training. The protein encoded by *PRDM1* is widely expressed and has been implicated in skeletal muscle fibre type differentiation (Beermann, Ardelt, Girgenrath, & Miller, 2010). The biology explaining why this gene variant is associated with $VO_{2\ max}$ has yet to be elucidated but, again in combination with other variants such as those identified by Bouchard et al. (2011), could provide a useful tool in rugby.

Additionally, the *ACTN3 R577X* gene variant (already mentioned) could, when combined with other genetic variants, be a molecular marker for muscle fibre type proportion. Vincent et al. (2007) showed that the RR genotype (presence of the protein) had ~5% greater type IIx fibres than XX genotype individuals (absence of the protein) and had ~5% larger relative surface area per fibre type ($n = 43$). Supporting evidence followed using a larger sample ($n = 94$) where the XX genotype was associated with ~3% greater type I muscle fibre type proportion than RR individuals (Ahmetov et al., 2011).

Injury risk

Possibly the most useful application of predictive genetics within sports genomics will be in the field of injury risk and severity estimation. Hakim et al. (2003) examined frozen shoulder (FS) and tennis elbow (TE) in 865 MZ and 963 DZ twin pairs and reported 42% heritability for FS and 40% for TE. Additionally, while we are not aware of heritability estimates for mild-to-moderate traumatic brain injury (TBI) such as concussion, the likelihood of suffering more severe future neurological disorders following these injuries is significant (Goldman et al., 2006).

Probably the most explored gene regarding tendon and ligament injuries is *COL5A1*, which encodes for the α1 chain of type V collagen, a minor fibrillar collagen (Hildebrand, Frank, & Hart, 2004). Mokone, Schwellnus, Noakes, and Collins (2006) investigated the difference in genotype frequency in South African Caucasians between 111 patients with Achilles tendon pathology and 129 control subjects. They showed that the CC genotype was over-represented in the control population, suggesting a protective role of the C allele against tendon injury.

A similar association, indicated a protective C allele, has been shown with cruciate ligament injuries in South African women (Posthumus et al., 2009). Assessment of this *COL5A1* genetic marker, in combination with other markers yet to be identified, might provide a useful tool in rugby for individualising training load and mode to reduce incidence of injury.

The *APOE* gene is on chromosome 19 and encodes apolipoprotein E-based peptide (ApoE), a candidate marker for risk and severity of TBI. ApoE is a protein that plays a pivotal role in cholesterol metabolism and has been linked to neurobiological function and, specifically, susceptibility to late onset and sporadic Alzheimer's disease via *APOE* gene polymorphisms (Bu, 2009). The *APOE* E4 allele has also been associated with neurodegenerative cascade subsequent to TBI and the severity of axonal injury in mouse models (Bennett et al., 2013), with human studies showing an association between the *APOE* E4 allele and poor outcome following TBI (Laskowitz & Vitek, 2007). Functionally, carriers of the *APOE* E4 allele have presented with reduced motor rehabilitation outcomes, poorer neurocognitive outcomes, increased cognitive impairments and memory defects following TBI (reviewed in Gokhale & Laskowitz, 2013). This genetic marker, combined with others such as *TNFA* −308, a promoter region within *GRIN2A*, and others including more yet to be identified, could improve estimation of both recovery duration from TBI/concussion in rugby and the risk of longer-term neurocongnitive problems – potentially useful from both short-term player management and long-term player health perspectives. The most interesting, but preliminary, investigation to date observed the recovery time of 51 athletes following medically diagnosed concussion; McDevitt et al. (2014) investigated a repeat polymorphism in the promoter region within the *GRIN2A* gene and found the chance of prolonged recovery from concussion was 4.3 times greater for homozygous risk allele carriers.

Skill acquisition and cognitive ability

In a large study of 11,000 twin pairs from four countries (USA, AUS, UK and The Netherlands), the heritability of general cognitive ability was found to increase linearly from childhood (41%) to adulthood (66%; Haworth et al., 2010), while the ability to learn motor skills is also highly heritable at ~70% (Fox, Hershberger, & Bouchard, 1996). One candidate to explain some of this heritability is the *BDNF* gene, which encodes for brain-derived neurotrophic factor, a protein that influences cortical synaptic plasticity (Akaneya, Tsumoto, Kinoshita, & Hatanaka, 1997). Individuals possessing the Met allele of the Val66Met polymorphism in the *BDNF* gene show lower increases in motor evoked potentials after motor training (Kleim et al., 2006). Furthermore, carriers of the Met allele showed no change in neurological excitability from transcranial magnetic stimulation, whereas carriers of the Val allele showed a 67% increase (Missitzi et al., 2011). These data suggest that carriers of the Met allele might show lower adaptation to motor learning through a lack of neurobiological excitability, possibly related to altered cortical synaptic plasticity. Another candidate is the Val158Met polymorphism in the catechol-*O*-methyl transferase (*COMT*) gene that encodes the COMT enzyme which is a major catabolising enzyme of the dopamine pathway. Individuals homozygous for the Met allele show impaired inhibition of prepotent responses (i.e. the ability to supress automated behavioural responses) and reduced working memory plasticity (Bellander et al., 2014). It would be fascinating to test the hypothesis that favourable alleles of these *BDNF* and *COMT* gene variations, and others related to neurobiological function, may be found more frequently in playing positions with greater requirement to execute relatively fine motor skills and greater responsibility for making tactical decisions in rugby such as the half backs. Genetic tests for skill acquisition and cognitive ability might then become useful additions to other non-genetic assessments in talent identification and individualisation of skill and tactical elements of training programmes.

Athlete status

De Moor et al. (2007) investigated the genetic component of athlete status in female 1000 DZ and 793 MZ twins and reported a heritability estimate of ~70% for athlete status. We are not aware of similar data for males. In a review published in 2012, at least 79 genetic markers were identified that had been associated, in at least one prior research paper, with elite athlete status (Ahmetov & Fedotovskaya, 2012). However, that number was reduced to 20 when the criterion was at least two prior research papers, and probably even some of those associations will not prove to be true as more data are accumulated. The first scientific investigation to assess the molecular genetic component of elite athlete status (Gayagay et al., 1998) showed a significant association of the *ACE* I/D variant (mentioned previously) with elite status in 64 rowers. Subsequently, the *ACE* I allele has been associated with elite performance in a variety of sports, though not consistently, and the research is probably best summarised by reference to the meta-analysis of Ma et al. (2013) who found that

the II genotype of *ACE* I/D was associated with physical performance (odds ratio 1.23), especially endurance performance (odds ratio 1.35).

For the *ACTN3* R577X variant already mentioned, the R allele has been consistently associated with elite power and sprint athletes from a variety of backgrounds and in top elite sprinters, a complete absence of the XX genotype has been identified (Niemi & Majamaa, 2005; Yang et al., 2003), compared to an XX genotype frequency of ~18% in the general Caucasian population. Again, the meta-analysis by Ma et al. (2013) nicely summarises the association of the R allele with elite power athlete status (odds ratio 1.21).

Gradually, a number of athlete cohorts have emerged, hosting steadily larger samples of elite and sub-elite athletes from various sports for the investigation of athlete status (Pitsiladis et al., 2013). We are also aware that a UK athlete cohort is being established. One potential application of knowledge about the genetic characteristics of elite athletes, including in rugby union, could be an enhancement to talent identification programmes that are in some cases already quite sophisticated.

Genetics and rugby union

As early as 1922 scientists were hypothesising about the role of heritability in rugby union athletes. Jack (1922) documented the playing positions in 23 sets of elite rugby-playing brothers (n = 63), including a number of international representatives, and concluded that "the ability required for playing in certain positions in rugby football is inherited". We are aware of only three full publications in peer-reviewed journals applying molecular genetics to rugby union. Goh et al. (2009) reported that the II *ACE* genotype was associated with a higher ventilatory threshold in non-elite Asian rugby players but the very small cohort (n = 17) is a major limitation to that study. Bell et al. recently assessed *ACE* I/D and *ACTN3* R577X genotypes in 68 and 102 young non-elite rugby union players, respectively (Bell, Colley, Evans, Darlington, & Cooper, 2012; Bell et al., 2010). No associations were identified between either polymorphism and athlete status, playing position or the physiological and anthropometric parameters assessed, perhaps due to the rather small sample size and the sub-elite status of the players. With such little data regarding the genetic characteristics of rugby union players in the literature, this gap should be filled with high quality data using appropriate experimental designs.

Next steps for rugby genomics research

Cohort

For sports genomics research to be truly relevant to the preparation and management of elite competitors, the athletes from which the initial data are derived must themselves be considered elite. In the context of rugby union, we propose a definition of "elite" as athletes competing in the highest competitive league of a "Tier 1" rugby nation. Comparisons between sub-elite (e.g. the next tier in the league structure), suitably matched controls and elite athletes would also be useful for explanatory purposes. Given the evolving nature of elite rugby, the era in which athletes competed at an elite level also needs to be defined. Rugby union has changed dramatically in the >100 years of its existence and that change has certainly continued even since the sport turned professional ~20 years ago. Nevertheless, we propose that 1995 onwards is a playing era inclusion criterion that can sensibly be justified.

Another critical characteristic of the cohort is that it is large in number. Genetic research requires a large sample size to obtain sufficient statistical power and simultaneously minimise the likelihood of Type 1 statistical error (false positive associations). Only by using a large cohort can patterns and associations in the data be identified reliably and sound conclusions made regarding associations of genetic variations within a given cohort. Consequently, large numbers (many hundreds, and preferably thousands) of "Tier 1" rugby athletes are required – and that will only be achieved through international research collaboration. Additionally, geographic ancestry is an important consideration for case–control and genotype–phenotype association study designs and therefore analysis of molecular genetic markers should preferably be performed on athletes from a well-defined geographic ancestry cluster. A more difficult aim to achieve would be to recruit large numbers of players from all geographic ancestry clusters commonly found in rugby union, although this would be a very powerful approach scientifically.

Genotype–phenotype associations that should be investigated

Following the recruitment of the appropriate cohort, the next step would be to investigate genotype associations with a range of phenotypes such as success in rugby (elite athlete status and playing position), physiological, anthropometric and other performance variables, incidence of injury, etc. We recently presented preliminary results involving 272 elite (as defined above) rugby union athletes (compared to 141 controls matched for geographic

ancestry) at a scientific conference, focusing on elite athlete status in that report. Specifically, we reported an association between the *ACTN3* R577X gene variant and elite rugby union athlete status as well as playing position (Heffernan et al., 2014). We are continuing to build this cohort in what we have termed the RugbyGene project.

Identifying genetic associations (in a cohort of elite rugby athletes) with rugby-specific physiological and anthropometric variables for example those aspects of strength, speed and body composition assessed by Smart et al. (2013), would further exemplify the importance of the genetic component to rugby. Furthermore, rugby union has one of the highest reported incidents of match play injuries in all professional team sports (Brooks & Kemp, 2008), with an injury incidence of 81 per 1000 match play hours resulting in ~20 days absence per injury (Williams, Trewartha, Kemp, & Stokes, 2013). Research collaborations that combine these kinds of large, meticulously collected rugby injury databases with genetic analyses conducted on those very same players, could be extremely fruitful in explaining some of the as yet unexplained inter-individual variability in injury susceptibility and could identify novel markers of injury risk in rugby.

Concussion risk in rugby, as well as consideration of the potential longer-term consequences, is clearly and justifiably a topic of much attention at this time (Raftery, 2014). Accordingly, Gardner et al. (2014) recently conducted a comprehensive meta-analysis showing that in men's rugby union, the incidence rate of concussion was ~4.7 per 1000 match play hours. Data from 16 studies showed that at the elite/international level the incidence rate was lower (~1.2 per 1000 match play hours), though still considerable (Gardner et al., 2014). These notable injury rates make research efforts to identify molecular markers for the risk of specific rugby-related injuries, such as concussion, highly warranted. Indeed, the development of tools to identify individuals at greater risk of concussion in rugby and greater risk of longer-term pathological neurobiological changes following a career playing rugby would seem highly responsible in the context of player welfare.

Advanced genomics technology

For rugby union as in other areas within sports genomics, the starting point for the investigation of genetic variation is through the candidate gene (hypothesis driven) approach including the genes mentioned here in this manuscript and many others. This approach is the process of considering the biological mechanisms of a given trait (for example, incidence of tendinopathy or high $VO_{2\ max}$) and investigating previously identified genetic variants within genes known or suspected to affect the relevant biological pathways in some way. While the candidate gene approach is by far the most utilised technique in quantifying molecular genetic markers of sport-related phenotypes and a good starting point, other more complex analysis techniques are required for this field to realise its full potential (Pitsiladis et al., 2013).

The hypothesis-free approach of conducting genome wide association studies (GWAS) has been utilised extensively to identify new genetic variants in various domains within human biology (Wolfarth et al., 2014) and is recommended for identifying novel genetic variants in rugby union. GWAS is the process of investigating large numbers of known SNPs simultaneously (~2 million, for example) for a given complex trait (Visscher, Brown, McCarthy, & Yang, 2012). As already mentioned, complex traits of relevance to rugby union could be sprinting ability, muscle strength, incidence of tendinopathy or simply being an elite rugby union player (elite athlete status). Importantly, given the large number of hypotheses tested statistically, only the strongest associations are usually accepted to be true results (e.g. when $P < 5 \times 10^{-8}$), although a very large cohort size and/or strong genetic effect sizes are usually required for this approach to be effective. A strength of any hypothesis-free approach like GWAS is that new variants which reveal new biological insight can be discovered – and then further investigated experimentally.

Genetic testing technologies have advanced to such an extent that investigating all nucleotides in a gene, all protein-coding genes, or even the whole genome is now possible via direct sequencing. Depending on the availability of funding, participants and other resources including laboratory equipment and bioinformatics expertise, multiple options are available to exploit these techniques for the identification of novel mutations, polymorphisms or structural variants. First, targeting specific genes associated with a given phenotype, such as the *MSTN* gene and muscle mass, and sequencing every nucleotide (~7000 bp) of that gene in a large cohort of elite rugby union players for whom strength and muscle mass are also known could be an elegant approach to identifying novel genetic variants associated with muscle size and strength in rugby players. To examine larger regions than one gene, one could utilise whole-exome sequencing, which uses the same rational as targeted gene sequencing but to a much greater extent, targeting all protein-coding DNA sequences (~230,000 exons or ~30 million bp; ~1% of the human genome). Finally, providing that considerable resources including an appropriately large cohort were available, assessment of the whole genome (~3 billion bp) would be the

ultimate, most comprehensive method of identifying novel mutations or polymorphisms of functional importance in rugby union athletes. Currently, whole-genome sequencing is not recommended for reasons of cost, logistics, interpretation and statistics, although the eventual application to rugby is inevitable.

Following the identification of novel gene variations through the methods described above, simple mathematical models can be applied to simultaneously quantify the effect of several gene variants. Williams and Folland (2008) developed the total genotype score (TGS) as a way to represent the simultaneous influence of multiple genetic markers as a simple value that is intuitively understandable. According to the TGS, all scores lie between 0 and 100 and a higher score indicates a greater genetic suitability for a given phenotype such as muscle mass, sprinting speed, reduced injury risk, greater cognitive ability – depending on the genetic variants that are used to calculate the specific TGS in each case. Thus, multiple TGS scores, each directed towards a given phenotype, could be calculated for each individual. Subsequently, both hypothetical (Hughes, Day, Ahmetov, & Williams, 2011) and experimental (Eynon, Ruiz, Meckel, Morán, & Lucia, 2011; Ruiz et al., 2009, 2010; Santiago et al., 2010) studies have used this method in attempts to elucidate the "optimal polygenetic profile" for a given cohort with some limited success. Other statistical methods to capture the simultaneous influence of multiple genetic markers include construction of inferred pseudo-haplotypes when appropriate or simply calculating the number of favourable alleles carried by an individual across a number of unrelated genetic loci.

Conclusion

To generate useful knowledge about the genomics of rugby union, hundreds or thousands of individual players are required in a research cohort for one to be confident that the data are robust – particularly when performing some of the more complex genomic analyses outlined in the preceding section. We and others are collaborating in a large, multi-institutional effort to make progress in this field – the RugbyGene project. There will need to be a simultaneous analysis of genomic and phenotypic (performance and injury) data to maximise the potential of this kind of research. The challenge will then be to apply genomic technologies, alongside existing non-genomic data of course, to personalise the management of players in elite rugby union and facilitate the prescription of training, nutrition, playing load and management of injury risk in a more individualised manner than is currently possible to improve both performance and player welfare.

Acknowledgements

The authors would like to acknowledge Matthew Dimbylow of Dream it, Believe it Achieve it (DBA) for financial support.

References

Ahmetov, I. I., Druzhevskaya, A. M., Lyubaeva, E. V., Popov, D. V., Vinogradova, O. L., & Williams, A. G. (2011). The dependence of preferred competitive racing distance on muscle fibre type composition and ACTN3 genotype in speed skaters. *Experimental Physiology*, *96*, 1302–1310. doi:10.1113/expphysiol.2011.060293

Ahmetov, I. I., & Fedotovskaya, O. N. (2012). Sports genomics: Current state of knowledge and future directions. *Cellular and Molecular Exercise Physiology*, *1*(1), e1. doi:10.7457/cmep.v1i1.e1

Akaneya, Y., Tsumoto, T., Kinoshita, S., & Hatanaka, H. (1997). Brain-derived neurotrophic factor enhances long-term potentiation in rat visual cortex. *The Journal of Neuroscience*, *17*, 6707–6716.

Almeida, S. S., Barros, C. C., Moraes, M. R., Russo, F. J., Haro, A. S., Rosa, T. S., … Araujo, R. C. (2010). Plasma kallikrein and angiotensin I-converting enzyme N- and C-terminal domain activities are modulated by the insertion/deletion polymorphism. *Neuropeptides*, *44*(2), 139–143.

Barr, M. J., Sheppard, J. M., Gabbett, T. J., & Newton, R. U. (2014). Long-term training induced changes in sprinting speed and sprint momentum in elite rugby union players. *Journal of Strength and Conditioning Research*, *28*, 2724–2731. doi:10.1519/JSC.0000000000000364

Beermann, M. L., Ardelt, M., Girgenrath, M., & Miller, J. B. (2010). Prdm1 (Blimp-1) and the expression of fast and slow myosin heavy chain isoforms during avian myogenesis in vitro. *PLoS ONE*, *5*, e9951. doi:10.1371/journal.pone.0009951

Bell, W., Colley, J. P., Evans, W. D., Darlington, S. E., & Cooper, S.-M. (2012). ACTN3 genotypes of rugby union players: Distribution, power output and body composition. *Annals of Human Biology*, *39*(1), 19–27. doi:10.3109/03014460.2011.632648

Bell, W., Colley, J. P., Gwynne, J. R., Glazier, P., Evans, W. D., & Darlington, S. E. (2010). ACE ID genotype and leg power in rugby union players. *Journal of Sports Medicine and Physical Fitness*, *50*, 350–355.

Bellander, M., Bäckman, L., Liu, T., Schjeide, B. M., Bertram, L., Schmiedek, F., … Lövdén, M. (2014). Lower baseline performance but greater plasticity of working memory for carriers of the val allele of the comt val158met polymorphism. *Neuropsychology*, *29*(2), 247–254. doi:10.1037/neu0000088

Bennett, R. E., Esparza, T. J., Lewis, H. A., Kim, E., Mac Donald, C. L., Sullivan, P. M., & Brody, D. L. (2013). Human apolipoprotein E4 worsens acute axonal pathology but not amyloid-β immunoreactivity after traumatic brain injury in 3xTG-AD mice. *Journal of Neuropathology & Experimental Neurology*, *72*, 396–403. doi:10.1097/NEN.0b013e31828e24ab

Bouchard, C., An, P., Rice, T., Skinner, J. S., Wilmore, J. H., Gagnon, J., … Rao, D. C. (1999). Familial aggregation of $VO_{2\,max}$ response to exercise training: Results from the HERITAGE family study. *Journal of Applied Physiology*, *87*, 1003–1008.

Bouchard, C., Sarzynski, M. A., Rice, T. K., Kraus, W. E., Church, T. S., Sung, Y. J., … Rankinen, T. (2011). Genomic predictors of the maximal O_2 uptake response to standardized

exercise training programs. *Journal of Applied Physiology, 110*, 1160–1170. doi:10.1152/japplphysiol.00973.2010

Bradley, W., Cavanagh, B., Douglas, W., Donovan, T. F., Morton, J. P., & Close, G. L. (2014). Quantification of training load, energy intake and physiological adaptations during a rugby preseason: A case study from an elite European rugby union squad. *Journal of Strength Conditioning Research, 29*(2), 534–544. doi:10.1519/jsc.0000000000000631

Brooks, J. H. M., & Kemp, S. (2008). Recent trends in rugby union injuries. *Clinics in Sports Medicine, 27*(1), 51–73. doi:10.1016/j.csm.2007.09.001

Bu, G. (2009). Apolipoprotein E and its receptors in Alzheimer's disease: Pathways, pathogenesis and therapy. *Nature Reviews Neuroscience, 10*, 333–344. doi:10.1038/nrn2620

Cahill, N., Lamb, K., Worsfold, P., Headey, R., & Murray, S. (2012). The movement characteristics of English premiership rugby union players. *Journal of Sports Sciences, 31*, 229–237. doi:10.1080/0264041 4.2012.727456

De Moor, M. H., Spector, T. D., Cherkas, L. F., Falchi, M., Hottenga, J. J., Boomsma, D. I., & De Geus, E. J. (2007). Genome-wide linkage scan for athlete status in 700 British female DZ twin pairs. *Twin Research and Human Genetics, 10*, 812–820.

ENCODE Project Consortium. (2012). An integrated encyclopedia of DNA elements in the human genome. *Nature, 489*, 57–74.

Eynon, N., Hanson, E. D., Lucia, A., Houweling, P. J., Garton, F., North, K. N., & Bishop, D. J. (2013). Genes for elite power and sprint performance: ACTN3 leads the way. *Sports Medicine, 43*, 803–817.

Eynon, N., Ruiz, J. R., Meckel, Y., Morán, M., & Lucia, A. (2011). Mitochondrial biogenesis related endurance genotype score and sports performance in athletes. *Mitochondrion, 11*(1), 64–69. doi:10.1016/j.mito.2010.07.004

Fox, P. W., Hershberger, S. L., & Bouchard, T. J. (1996). Genetic and environmental contributions to the acquisition of a motor skill. *Nature, 384*, 356–358. doi:10.1038/384356a0

Frazer, K. A., Ballinger, D. G., Cox, D. R., Hinds, D. A., Stuve, L. L., Gibbs, R. A., ... Leal, S. M. (2007). A second generation human haplotype map of over 3.1 million SNPs. *Nature, 449*, 851–861.

Fuller, C. W., Taylor, A. E., Brooks, J. H., & Kemp, S. P. (2013). Changes in the stature, body mass and age of English professional rugby players: A 10-year review. *Journal of Sports Sciences, 31*, 795–802. doi:10.1080/02640414.2012.753156

Gardner, A. J., Iverson, G. L., Williams, W. H., Baker, S., & Stanwell, P. (2014). A systematic review and meta-analysis of concussion in rugby union. *Sports Medicine, 12*, 1–15. doi:10.1007/s40279-014-0233-3

Gayagay, G., Yu, B., Hambly, B., Boston, T., Hahn, A., Celermajer, D. S., & Trent, R. J. (1998). Elite endurance athletes and the ACE I allele – The role of genes in athletic performance. *Human Genetics, 103*(1), 48–50.

Goh, K. P., Chew, K., Koh, A., Guan, M., Wong, Y. S., & Sum, C. F. (2009). The relationship between ACE gene ID polymorphism and aerobic capacity in Asian rugby players. *Singapore Medical Journal, 50*, 997–1003.

Gokhale, S., & Laskowitz, D. T. (2013). ApoE and outcome after traumatic brain injury. *Clinical Lipidology, 8*, 561–571.

Goldman, S. M., Tanner, C. M., Oakes, D., Bhudhikanok, G. S., Gupta, A., & Langston, J. W. (2006). Head injury and Parkinson's disease risk in twins. *Annals of Neurology, 60*(1), 65–72. doi:10.1002/ana.20882

Hakim, A. J., Cherkas, L. F., Spector, T. D., & MacGregor, A. J. (2003). Genetic associations between frozen shoulder and tennis elbow: A female twin study. *Rheumatology, 42*, 739–742.

Haworth, C. M., Wright, M. J., Luciano, M., Martin, N. G., De Geus, E. J., Van Beijsterveldt, C. E., ... Davis, O. S. (2010). The heritability of general cognitive ability increases linearly from childhood to young adulthood. *Molecular Psychiatry, 15*, 1112–1120.

Heffernan, S. M., Kilduff, L. P. M., Erskine, R., Day, S. H., McMahon, G. E., Cook, C. J., ... Williams, A. G. (2014). ACE and ACTN3 gene variants are associated with elite status and playing position in rugby union. In Bases 2013 abstracts. *Journal of Sports Sciences, 32*(Suppl. 1), s49–s50. doi:10.1080/02640414.2014.896604

Hildebrand, K. A., Frank, C. B., & Hart, D. A. (2004). Gene intervention in ligament and tendon: Current status, challenges, future directions. *Gene Therapy, 11*, 368–378. doi:10.1038/sj.gt.3302198

Hughes, D. C., Day, S. H., Ahmetov, I. I., & Williams, A. G. (2011). Genetics of muscle strength and power: Polygenic profile similarity limits skeletal muscle performance. *Journal of Sports Sciences, 29*, 1425–1434.

Jack, H. W. (1922). Brothers in rugby football: More evidence that brothers are selected to play in the same or similar positions on football teams. *Journal of Heredity, 13*, 161–162.

Kleim, J. A., Chan, S., Pringle, E., Schallert, K., Procaccio, V., Jimenez, R., & Cramer, S. C. (2006). BDNF val66met polymorphism is associated with modified experience-dependent plasticity in human motor cortex. *Nature Neuroscience, 9*, 735–737.

Laskowitz, D. T., & Vitek, M. P. (2007). Apolipoprotein E and neurological disease: Therapeutic potential and pharmacogenomic interactions. *Pharmacogenomics, 8*, 959–969.

Liu, X. G., Tan, L. J., Lei, S. F., Liu, Y.-J., Shen, H., Wang, L., ... Chen, X. D. (2009). Genome-wide association and replication studies identified TRHR as an important gene for lean body mass. *The American Journal of Human Genetics, 84*, 418–423.

Ma, F., Yang, Y., Li, X., Zhou, F., Gao, C., Li, M., & Gao, L. (2013). The association of sport performance with ACE and ACTN3 genetic polymorphisms: A systematic review and meta-analysis. *PLoS ONE, 8*, e54685.

McDevitt, J., Tierney, R., Phillips, J., Gaughan, J., Torg, J. S., & Krynetskiy, E. (2014). Supporting the concept of genetic predisposition to prolonged recovery following a concussion. *Orthopaedic Journal of Sports Medicine, 2*(Suppl. 2), 1–2. doi:10.1177/2325967114S00078

Missitzi, J., Gentner, R., Geladas, N., Politis, P., Karandreas, N., Classen, J., & Klissouras, V. (2011). Plasticity in human motor cortex is in part genetically determined. *The Journal of Physiology, 589*, 297–306.

Mokone, G. G., Schwellnus, M. P., Noakes, T. D., & Collins, M. (2006). The COL5A1 gene and achilles tendon pathology. *Scandinavian Journal of Medicine & Science in Sports, 16*(1), 19–26.

Niemi, A.-K., & Majamaa, K. (2005). Mitochondrial DNA and ACTN3 genotypes in Finnish elite endurance and sprint athletes. *European Journal of Human Genetics, 13*, 965–969.

North, K. N., Yang, N., Wattanasirichaigoon, D., Mills, M., Easteal, S., & Beggs, A. H. (1999). A common nonsense mutation results in α-actinin-3 deficiency in the general population. *Nature Genetics, 21*, 353–354.

Peeters, M. W., Thomis, M. A., Loos, R. J. F., Derom, C. A., Pagard, R., Claessens, A. L., ... Beunen, G. P. (2007). Heritability of somatotype components: A multivariate analysis. *International Journal of Obesity, 31*(8), 1295–1301.

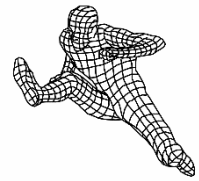
Pitsiladis, Y., Wang, G., Wolfarth, B., Scott, R., Fuku, N., Mikami, E., ... Lucia, A. (2013). Genomics of elite sporting performance: What little we know and necessary advances. *British Journal of Sports Medicine, 47*, 550–555.

Posthumus, M., September, A. V., Keegan, M., O'Cuinneagain, D., Van der Merwe, W., Schwellnus, M. P., & Collins, M. (2009). Genetic risk factors for anterior cruciate ligament

ruptures: COL1A1 gene variant. *British Journal of Sports Medicine, 43*, 352–356.
Puthucheary, Z., Skipworth, J. R. A., Rawal, J., Loosemore, M., Van Someren, K., & Montgomery, H. E. (2011). The ACE gene and human performance. *Sports Medicine, 41*, 433–448.
Quarrie, K. L., Hopkins, W. G., Anthony, M. J., & Gill, N. D. (2012). Positional demands of international rugby union: Evaluation of player actions and movements. *Journal of Science and Medicine in Sport, 16*, 353–359.
Raftery, M. (2014). Concussion and chronic traumatic encephalopathy: International Rugby Board's response. *British Journal of Sports Medicine, 48*(2), 79–80.
Roberts, S. P., Trewartha, G., Higgitt, R. J., El-Abd, J., & Stokes, K. A. (2008). The physical demands of elite English rugby union. *Journal of Sports Sciences, 26*, 825–833.
Ruiz, J. R., Arteta, D., Buxens, A., Artieda, M., Gómez-Gallego, F., Santiago, C., … Lucia, A. (2010). Can we identify a power-oriented polygenic profile? *Journal of Applied Physiology, 108*, 561–566.
Ruiz, J. R., Gómez-Gallego, F., Santiago, C., González-Freire, M., Verde, Z., Foster, C., & Lucia, A. (2009). Is there an optimum endurance polygenic profile? *The Journal of Physiology, 587*, 1527–1534.
Sachidanandam, R., Weissman, D., Schmidt, S. C., Kakol, J. M., Stein, L. D., Marth, G., … International SNP Map Working Group. (2001). A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. *Nature, 409*, 928–933.
Santiago, C., Ruiz, J. R., Muniesa, C. A., González-Freire, M., Gómez-Gallego, F., & Lucia, A. (2010). Does the polygenic profile determine the potential for becoming a world-class athlete? Insights from the sport of rowing. *Scandinavian Journal of Medicine and Science in Sports, 20*(1), e188–e194.
Schuelke, M., Wagner, K. R., Stolz, L. E., Hübner, C., Riebel, T., Kömen, W., … Lee, S.-J. (2004). Myostatin mutation associated with gross muscle hypertrophy in a child. *New England Journal of Medicine, 350*, 2682–2688.
Silventoinen, K., Magnusson, P. K., Tynelius, P., Kaprio, J., & Rasmussen, F. (2008). Heritability of body size and muscle strength in young adulthood: A study of one million Swedish men. *Genetic Epidemiology, 32*, 341–349.
Simoneau, J.-A., & Bouchard, C. (1995). Genetic determinism of fiber type proportion in human skeletal muscle. *The FASEB Journal, 9*, 1091–1095.
Smart, D., Hopkins, W. G., Quarrie, K. L., & Gill, N. (2011). The relationship between physical fitness and game behaviours in rugby union players. *European Journal of Sport Science, 14*, 1–10. doi:10.1080/17461391.2011.635812
Smart, D. J., Hopkins, W. G., & Gill, N. D. (2013). Differences and changes in the physical characteristics of professional and amateur rugby union players. *The Journal of Strength & Conditioning Research, 27*, 3033–3044.
The International HapMap Consortium. (2005). A haplotype map of the human genome. *Nature, 437*, 1299–1320.
Vincent, B., De Bock, K., Ramaekers, M., Van den Eede, E., Van Leemputte, M., Hespel, P., & Thomis, M. A. (2007). ACTN3 (R577X) genotype is associated with fiber type distribution. *Physiological Genomics, 32*(1), 58–63.
Visscher, P. M., Brown, M. A., McCarthy, M. I., & Yang, J. (2012). Five years of GWAS discovery. *The American Journal of Human Genetics, 90*(1), 7–24.
West, D., Cunningham, D., Bevan, H., Crewther, B., Cook, C., & Kilduff, L. (2013). Influence of active recovery on professional rugby union player's ability to harness postactivation potentiation. *The Journal of Sports Medicine and Physical Fitness, 53*, 203–208.
Williams, A. G., & Folland, J. P. (2008). Similarity of polygenic profiles limits the potential for elite human physical performance. *The Journal of Physiology, 586*(1), 113–121.
Williams, S., Trewartha, G., Kemp, S., & Stokes, K. (2013). A meta-analysis of injuries in senior men's professional rugby union. *Sports Medicine, 43*, 1043–1055.
Wolfarth, B., Rankinen, T., Hagberg, J. M., Loos, R. J., Pérusse, L., Roth, S. M., … Bouchard, C. (2014). Advances in exercise, fitness, and performance genomics in 2013. *Medicine and Science in Sports and Exercise, 46*, 851–859.
Yang, N., MacArthur, D. G., Gulbin, J. P., Hahn, A. G., Beggs, A. H., Easteal, S., & North, K. (2003). ACTN3 genotype is associated with human elite athletic performance. *The American Journal of Human Genetics, 73*(3), 627–631.

Athlete participation pack

Manchester Metropolitan University — Cheshire —



Department of Exercise and Sport Science

Informed Consent Form

(Both the investigator and participant should retain a copy of this form)

Name of Participant:

Principal Investigator: Dr Alun Williams

Project Title: The Genetic Profile of Elite Athletes

Ethics Committee Approval Number: 12.07.11 (i)

Participant Statement

I have read the participant information sheet for this study and understand what is involved in taking part. Any questions I have about the study, or my participation in it, have been answered to my satisfaction. I understand that I do not have to take part and that I may decide to withdraw from the study at any point without giving a reason. Any concerns I have raised regarding this study have been answered and I understand that any further concerns that arise during the time of the study will be addressed by the investigator. I therefore agree to participate in the study.

It has been made clear to me that, should I feel that my rights are being infringed or that my interests are otherwise being ignored, neglected or denied, I should inform the The University Secretary and Clerk to the Board of Governors, Manchester Metropolitan University, Ormond Building, Manchester, M15 6BX. Tel: 0161 247 3400 who will undertake to investigate my complaint.

Signed (Participant)		Date	
Signed (Investigator)		Date	

The Genetic Profile of Elite Athletes: Questionnaire

Thank you for your interest in our research study. Please answer the following questions about your ethnic origin, athlete status, and your training, diet and injury history.

SECTION A: Questions concerning your ethnic background.

Participant ID code: _____ Date of birth: _____

Gender (please tick): Male ☐ / Female ☐ Height (in metres): _____

Nationality (as on passport, e.g. British): _____ Body weight (in kg): _____

What is your ethnic group? Please tick the appropriate box.

A) White: English ☐ Scottish ☐ Welsh ☐ N. Irish ☐ Irish ☐
French ☐ South African ☐ New Zealander ☐ Australian ☐ Other ☐
If other, please state here: _____

B) Mixed: White & Black British ☐ White & Black Caribbean ☐ White & Black African ☐ White & Asian ☐ White & Latin American ☐ Other ☐
If other, please state here: _____

C) Asian: British ☐ Indian ☐ Pakistani ☐ Chinese ☐ Japanese ☐ Other ☐
If other, please state here: _____

D) Black: British ☐ Caribbean ☐ African ☐ Other ☐
If other, please state here: _____

E) Latin American: Brazilian ☐ Argentinian ☐ Mexican ☐ Colombian ☐ Other ☐
If other, please state here: _____

F) Pacific Islands: Samoa ☐ Fiji ☐ Tonga ☐ PNG ☐ Other ☐
If other, please state here: _____

G) Other ethnic background: ☐ Please state here: _____

I do not wish to state my ethnic origin ☐

Using the ethnic groups above as a guide, please tell us the ethnic origin of your:

Mother: _____ Don't know: ☐

Father: _____ Don't know: ☐

Mother's mother: _____ Don't know: ☐

Mother's father: _____ Don't know: ☐

Father's mother: _____ Don't know: ☐

Father's father: _____ Don't know: ☐

Blood donation

We would like to take a small (10 mL) blood sample from a vein in your arm. Before doing so, please answer the following safety questions.

1. Have you ever been infected with a blood-borne disease? _____ Yes ☐ No ☐
2. Are you anaemic or receiving treatment for anaemia or iron deficiency? __ Yes ☐ No ☐

If you have answered YES to any of these questions and/or you would prefer not to provide a blood sample, a saliva sample may be provided instead.

PLEASE TURN OVER

SECTION B: Questions concerning your athlete status.

1. What is/was your main playing position (if team sport) or your main event (if individual sport). If multiple, please state preferred position/event.
   _____
   _____

2. Please state the number of seasons you have competed as a professional: _____

3. Please state all the professional clubs you have competed for so far in your career:
   _____
   _____

4. Please state the highest level that you have competed, including number of caps earned e.g. England under 16s (4 caps), 18s (10 caps), senior (21 caps): _____
   _____

5. Have you any other athletic achievements? If so please state highest achievements and include relevant details:
   _____
   _____

SECTION C: Questions concerning your training.

1. Typically, how many hours do you train a week?

   _____

2. Typically, what is your average running distance per week?

   _____

SECTION D: Questions concerning your injury history.

10. Have you ever ruptured your tendon? Yes ☐ If yes, which tendon? *e.g. Achilles* No ☐

11. If Yes, please give details of how this occurred and at what age. **Activity** *e.g. sprinting* **Age** *e.g. 20*

12. Have you ever suffered from prolonged tendon pain during exercise that does not go away for weeks? Yes ☐ If yes, which tendon? *e.g. Achilles* No ☐

13. Have you ever been told that you have had tendinopathy? Yes ☐ If yes, which tendon? *e.g. Achilles* No ☐

14. If Yes, was it confirmed by a scan, e.g. MRI or ultrasound? Yes ☐ No ☐

15. Does anyone in your close family suffer from tendinopathy? Yes ☐ No ☐ Don't know ☐

If yes, which tendon?
e.g. Achilles

16. Has anyone in your close family ever ruptured a tendon? Yes ☐ No ☐ Don't know ☐

If yes, which tendon?
e.g. Achilles

17. Have you ever fully ruptured a ligament? Yes ☐ If yes, which ligament? *e.g. ACL* No ☐

18. If Yes, please give details of how this occurred and at what age.

Contact	**Non-contact**	**Age**
e.g. tackled from the side	*e.g. landing from a jump*	*e.g. 20*

19. Have you ever been told that you have had a ligament sprain/tear? Yes ☐ If yes, which ligament? *e.g. ACL* No ☐

20. If Yes, was it confirmed by a scan, e.g. MRI or ultrasound? Yes ☐ No ☐

21. Has anyone in your close family ever ruptured a ligament? Yes ☐ No ☐ Don't know ☐

If yes, which ligament? *e.g. ACL*

22. Have you ever been concussed or knocked out? Yes ☐ No ☐

23. If Yes, how many times have you been concussed or knocked out? _____ times

24. What were you doing at the time of the injury(ies)? E.g. rugby tackle, boxing, road accident. _____
_____

25. If Yes, how long was your recovery period, until the day when you had no signs and symptoms and were free to train and play fully? (tick, multiple times if necessary, any recovery periods that apply for the different occasions)
<7 days ☐ 7-10 days ☐ 10-20 days ☐
20-40 days ☐ 40-60 days ☐ >60 days ☐

26. If Yes, was/were your concussion(s) or knock-out(s) diagnosed by a medical professional? (tick, multiple times if necessary, any that apply) Yes ☐ No ☐

27. Does anyone in your close family (parents, siblings or grandparents) suffer from a neurological condition, such as:
Dementia, Alzheimer's disease, chronic traumatic encephalopathy (CTE), cognitive impairment, movement disorders, psychiatric disorders, motor neuron disease
Yes ☐ No ☐ Don't know ☐

Who and which condition(s)?
e.g. grandfather, dementia

_____

_____

Contact Details

Thank you for your interest in this research study. Please provide us with your contact details, so that we may contact you with information at a later date.

Name (PLEASE PRINT CLEARLY):
_____
Email (PLEASE PRINT CLEARLY):
_____
Telephone (ONLY REQUIRED IF NOT USING EMAIL):
_____
Postal address (ONLY REQUIRED IF NOT USING EMAIL):
_____
_____
__
_____
__

The ACTN3 R577X gene variant

We all have two copies of the *ACTN3* gene, one inherited from each parent. At a certain point along the length of the gene the structure can vary slightly, which means that a particular protein (alpha actinin-3) can/cannot be produced in the muscle. Each person is either RR, RX or XX genotype for the *ACTN3* R577X gene variant. If you are XX genotype, you cannot produce alpha actinin-3, which is only found in fast-twitch muscle fibres. As these muscle fibres are important for producing force and power during high-speed muscle contractions, not having the protein might be detrimental for power generation. RR genotype is generally associated with strength, power and greater muscle size, while XX is linked to elite endurance athlete status.

Are you interested in receiving feedback regarding your ACTN3 R577X genotype?

Yes ☐ No ☐

If yes, would you prefer to receive that feedback via email?

Yes ☐ No ☐

Thank you for taking part in this project. All information will be kept strictly confidential.

The Genetic Profile of Elite Athletes

Questionnaire: Physical Activity & General Health

Manchester Metropolitan University

Thank you for participating in this research study. We would like you to answer a few questions concerning your general health and physical activity level. Please answer the following questions as honestly as you can.

Participant ID code: _____ Date of birth: _____

Gender (please tick): Male ☐ / Female ☐ Height: _____

Nationality (as on passport, e.g. British): _____ Body weight: _____

What is your ethnic group? Please tick the appropriate box.

A) White: English ☐ Scottish ☐ Welsh ☐ N. Irish ☐ Irish ☐

French ☐ South African ☐ New Zealander ☐ Australian ☐ Other ☐

If other, please state here: _____

B) Mixed: White & Black British ☐ White & Black Caribbean ☐ White & Black African ☐ White & Asian ☐ White & Latin American ☐ Other ☐

If other, please state here: _____

C) Asian: British ☐ Indian ☐ Pakistani ☐ Chinese ☐ Japanese ☐ Other ☐

If other, please state here: _____

D) Black: British ☐ Caribbean ☐ African ☐ Other ☐

If other, please state here: _____

E) Latin American: Brazilian ☐ Argentinian ☐ Mexican ☐ Colombian ☐ Other ☐

If other, please state here: _____

F) Pacific Islands: Samoa ☐ Fiji ☐ Tonga ☐ PNG ☐ Other ☐

If other, please state here: _____

G) Other ethnic background: ☐ Please state here: _____

I do not wish to state my ethnic origin ☐

Using the ethnic groups above as a guide, please tell us the ethnic origin of your:

Mother: _____ Do not know: ☐

Father: _____ Do not know: ☐

Mother's mother: _____ Do not know: ☐

Mother's father: _____ Do not know: ☐

Father's mother: _____ Do not know: ☐

Father's father: _____ Do not know: ☐

Blood donation

We would like to take a small (10 mL) blood sample from a vein in your arm. Before doing so, please answer the following safety questions.

1. Have you ever been infected with a blood-borne disease? _____ Yes ☐ No ☐
2. Are you anaemic or receiving treatment for anaemia or iron deficiency? __ Yes ☐ No ☐

If you have answered YES to any of these questions and/or you would prefer not to provide a blood sample, a saliva sample may be provided instead.

Your general health

1. **At present,** do you have any health problem for which you are:

a) on medication, prescribed (by a doctor) or otherwise _____ Yes ☐ No ☐

b) attending (visiting) your doctor _____ Yes ☐ No ☐

c) on a hospital waiting list _____ Yes ☐ No ☐

2. **Have you ever** had any of the following?

a) Your doctor advised you not to take vigorous exercise _____ Yes ☐ No ☐

b) Pain in your chest when you undertake physical activity? _____ Yes ☐ No ☐

c) Central Nervous System disease, such as Parkinson, Alzheimer, Convulsions/epilepsy _____ Yes ☐ No ☐

d) Have you any history of chest problems, such as bronchitis, asthma or wheezy chest _____ Yes ☐ No ☐

e) Major illness, such as viral hepatitis, cancer _____ Yes ☐ No ☐

f) Eczema _____ Yes ☐ No ☐

g) Diabetes _____ Yes ☐ No ☐

h) High blood pressure _____ Yes ☐ No ☐

i) A limb fracture _____ Yes ☐ No ☐

j) Blood disorder, such as clotting problems, thrombosis, aneurysm, embolus) _____ Yes ☐ No ☐

k) Head injury _____ Yes ☐ No ☐

l) Digestive problems _____ Yes ☐ No ☐

m) Heart problems, such as heart attack, valve disease, palpitations, angina _____ Yes ☐ No ☐

n) Problems with bones, such as osteoporosis or osteoarthritis _____ Yes ☐ No ☐

o) Problems with joints, such as rheumatoid arthritis, any persistent pain, or any surgery on your joints _____ Yes ☐ No ☐

p) Back problems _____ Yes ☐ No ☐

q) Disturbance of balance/co-ordination, such as dizziness or balance-system dysfunction _____ Yes ☐ No ☐

r) Numbness in hands or feet _____ Yes ☐ No ☐

s) Disturbance of vision _____ Yes ☐ No ☐

t) Physical limitations, such as visual, hearing, walking problems ____ Yes ☐ No ☐

u) Thyroid problems, e.g. rapid loss or gain of weight _____ Yes ☐ No ☐

v) Kidney or liver problems _____ Yes ☐ No ☐

w) A severe allergic reaction, e.g. swelling, breathing difficulties in response to an external stimulus _____ Yes ☐ No ☐

x) Emotional or psychiatric problems _____ Yes ☐ No ☐

y) Any other illness or condition that affects your general health or interferes with your daily activities _____ Yes ☐ No ☐

4. If you answered **YES** to any of the questions above, please describe the details briefly below or to the investigator if you wish.

_____

_____

5. Are you currently involved in any other research studies at the University or elsewhere? Yes ☐ No ☐

If **YES** please provide details of the study:

_____

_____

Habitual physical activity

1. What is your main occupation? _____

2. At work I sit — Never ☐ Seldom ☐ Sometimes ☐ Often ☐ Always ☐

3. At work I stand — Never ☐ Seldom ☐ Sometimes ☐ Often ☐ Always ☐

4. At work I walk — Never ☐ Seldom ☐ Sometimes ☐ Often ☐ Always ☐

5. At work I lift heavy loads — Never ☐ Seldom ☐ Sometimes ☐ Often ☐ Always ☐

6. After work I am tired — Never ☐ Seldom ☐ Sometimes ☐ Often ☐ Always ☐

7. At work I sweat — Never ☐ Seldom ☐ Sometimes ☐ Often ☐ Always ☐

8. In comparison with others my own age I think my work is:

   Much heavier ☐ Heavier ☐ As heavy ☐ Lighter ☐ Much lighter ☐

9. Do you play sport or exercise? Yes ☐ No ☐

   If **YES**, which sport do you play most frequently? _____

How many hours per week?	Less than 1 ☐	1 to 2 ☐	2 to 3 ☐	3 to 4 ☐	More than 4 ☐
Time per session (hours)	½ ☐	1 ½ ☐	2 ½ ☐	3 ½ ☐	4 ½ ☐
How many months per year?	Less than 1 ☐	1 to 3 ☐	4 to 6 ☐	7 to 9 ☐	More than 9 ☐
What proportion of the month?	A few hours ☐	A few days ☐	2 weeks ☐	3 weeks ☐	Most of the month ☐

   If you do a **second** sport (or exercise class), which is it? _____

How many hours per week?	Less than 1 ☐	1 to 2 ☐	2 to 3 ☐	3 to 4 ☐	More than 4 ☐
Time per session (hours)	½ ☐	1 ½ ☐	2 ½ ☐	3 ½ ☐	4 ½ ☐
How many months per year?	Less than 1 ☐	1 to 3 ☐	4 to 6 ☐	7 to 9 ☐	More than 9 ☐
What proportion of the month?	A few hours ☐	A few days ☐	2 weeks	3 weeks	Most of the month ☐

☐ ☐

10. Compared with others of my own age I think my physical activity during leisure time is:

Much more ☐ More ☐ The same ☐ Less ☐ Much less ☐

11. During leisure time I sweat — Very Often ☐ Often ☐ Sometimes ☐ Seldom ☐ Never ☐

12. During leisure time I play sport — Never ☐ Seldom ☐ Sometimes ☐ Often ☐ Always ☐

13. During leisure time I watch TV — Never ☐ Seldom ☐ Sometimes ☐ Often ☐ Always ☐

14. During leisure time I walk — Never ☐ Seldom ☐ Sometimes ☐ Often ☐ Always ☐

15. During leisure time I cycle — Never ☐ Seldom ☐ Sometimes ☐ Often ☐ Always ☐

16. How many minutes do you walk per day to and from work, school and/or shopping?

Less than 5 ☐ 5 to 15 ☐ 16 to 30 ☐ 31 to 45 ☐ More than 45 ☐

Thank you for completing this questionnaire. All information will be kept strictly confidential.

Participant information sheet

This participant sheet includes information on additional testing procedures not conducted in the present thesis, but are performed as part of the larger GENESIS project (of which this thesis is part).

Manchester Metropolitan University
— Cheshire —

MANCHESTER METROPOLITAN UNIVERSITY
Department of Exercise and Sport Science

Information Sheet for Participants

Title of Study:
The Genetic Profile of Elite Athletes

Ethics Committee Reference Number: 12.07.11 (i)

1) This is an invitation to take part in a piece of research.
You are being invited to take part in a research study. Before you decide whether or not to take part, it is important for you to understand why the research is being done and what it will involve. Please take time to read the following information carefully and discuss it with others if you wish. Ask us if there is anything that is not clear or if you would like more information. Please take time to decide whether or not you wish to take part.

2) What is the purpose of the research?
The main purpose of the project is to investigate the influence of genetic differences on elite strength/power, mixed-demand and endurance athlete status compared to non-athletes.

3) Why is the study being performed?

Previous scientific studies have demonstrated that differences in genetic make-up are linked to elite athlete status, although relatively few gene variants have been identified and sometimes not enough participants have been recruited to find meaningful results. This study is being performed in very large groups of elite strength/power, mixed-demand and endurance athletes, as well as non-athletes, in order to learn more about which genes influence elite athlete status.

4) Why am I being asked to take part?
We are trying to recruit people from different populations, i.e. top-level strength/power, mixed-demand and endurance athletes, as well as non-athletes. This way we can compare the genetics of elite athletes with the general population.

5) Do I have to take part?

You are under no obligation to take part in this study. If, after reading this information sheet and asking any additional questions, you do not feel comfortable participating in the study you do not have to. If you do decide to take part, you are free to withdraw from the study at any point, without having to give a reason. If you do withdraw from the study you are free to take any personal data with you, on written request to the principal investigator, and this will not be included when the research is reported. If you decide not to take part or withdraw from the study it will not affect your relationship with any of the staff at the Manchester Metropolitan University.

If you do decide to take part you will be asked to sign an informed consent form stating your agreement to take part and you will be given a copy together with this information sheet to keep.

6) What will happen to me if I agree to take part?
In order to obtain the following information, you will either i) be visited once by one of the investigators, ii) asked to come to the Crewe campus of the University on one occasion, or iii) sent a pack, then asked to complete information sheets and simple DNA-collection procedures (saliva sample), and send them to the investigators. In the case of iii), you will be guided through the correct procedures during a telephone call made by one of the named members of the research team (please see below).

Questionnaires. We will ask you to complete a short survey that will give us an indication of your physical activity level, general health, ethnicity, sporting discipline and performance achievement, recent exercise and diet, injury history and one psychological trait questionnaire known as ‘mental toughness’. If you are female, you will also be asked about your menstruation status. The questionnaires will take about 30 minutes to complete.

Body height, weight and somatotype. We will measure your height and weight using standard equipment, from which we will be able to calculate your body mass index. This will take 5 minutes to complete. If you are comfortable doing so and if time permits, we will also determine your somatotype, which will involve taking skinfold measurements from multiple sites on your body (you will need to wear shorts and t-shirt/sports-bra). This will take a further 20 minutes to complete.

Blood sample/saliva sample. If you are visited by one of the investigators or you are asked to visit the University laboratory, you will be asked to provide a small blood sample, from which we will be able to analyse your DNA, RNA and protein levels. While you are lying down, a qualified phlebotomist will take 10 mL blood from a vein in your arm. This is a relatively painless procedure and will take less than a minute to complete. Alternatively, if blood sampling is not possible, you will be asked to dribble 2 mL of saliva into a tube for a few minutes and send the sample to the investigators. These procedures are completely harmless and painless and you will be guided through the correct procedures by one of the research team members.

Bone mineral density. You will be asked to undertake a bone ultrasound or peripheral quantitative computed tomography (pQCT) scan to give us information about your bone mineral density. The ultrasound scan involves a member of the research team scanning your shinbone with a small probe on your skin with the help of a lubricating gel for approximately 5 minutes. This procedure is completely harmless and painless. The pQCT scan (a bit like a medical X-ray) lasts approximately 10 minutes and involves your lower leg being placed in a supported, still position in the centre of the scanner. If you visit the university laboratory you will

be asked to complete a dual-energy X-ray absorptiometry (DEXA) scan which gives us information about your bone mineral density as well as muscle mass and fat mass. This procedure is also somewhat similar to that of a medical X-ray. You will be asked to lie on a plinth and remain as still as possible throughout the scan which will last approximately 8 minutes. Each DEXA and pQCT scan exposes you to an extremely minimal dose of radiation, which is well below the maximum recommended dose regarded as safe (see question 7 for more details).

Muscular strength and power. Muscular strength and power will measured using a cycle ergometer and a force platform. You will be asked to cycle as hard as you can for a short time (6 seconds) on the cycle ergometer. Using the force platform, you will be asked to do (i) a maximal jump and (ii) a maximal pull on a static bar while you are standing almost upright (knees slightly bent). You will be assisted by a member of the research team through all tests which will include warm-up and cool-down periods. To complete all strength and power tests will require up to 30 minutes.

7) Are there any disadvantages or risks in taking part?

Blood sample. This is not a painful procedure but some people are a little squeamish about blood and tend to faint. Therefore, you will be seated/lying down while we take the blood and you do not have to see anything. Sometimes there is a little bruising but this should disappear in a matter of a few days, even in the most extreme cases. To prevent further bleeding, you might be asked to place a cotton wool ball over the punctured site and to hold it in place for a few minutes or until bleeding has stopped. Blood samples will be stored in a locked freezer until analysis at a later date. Only appropriately qualified personnel will be used for taking the blood sample. *Any personal information provided by you in connection with the blood donation will be held in strict confidence. Furthermore, all data will be anonymised and stored in secure locations to prevent identification of an individual.*

DEXA/pQCT Scan. Should you decide to take part in this research, you will be exposed to a very small amount of radiation, specifically 8 µSv (DEXA) or 10 µSv (pQCT), depending on which scan is completed. This dose is extremely minimal and is equivalent to the amount of radiation you are subjected to over an average 2-3 days of your life or simply travelling 30-40 miles in a car.

Muscular strength and power. The three tests of muscle strength and power require short, maximal efforts, so you should not find them tiring. Like any physical effort there is a risk of muscle strain, but by ensuring you are fully warmed up this risk will be kept to a minimum.

8) What are the possible benefits of taking part?

The broad benefits of the research are linked to the potential for the study to highlight new links between gene variants and elite athlete status. This information will improve our understanding of what contributes to making an elite athlete ‘elite’ and which genes are responsible for determining the strength, size, power and endurance capacity of human muscle. By providing the necessary information, you will be contributing to our further understanding of how the body works and what makes us different from one another. Furthermore, we can provide you with immediate feedback concerning your bone mineral density and muscle strength and power, plus feedback concerning *ACTN3* genotype once the genetic analysis has taken place at a later date. This gene variant has been associated with elite strength/power and endurance athlete status.

9) Who are the members of the research team?

- Dr Alun Williams (MMU): Principal Investigator; responsible for overall project design and management, collecting blood and saliva samples, conducting other measurements and administering questionnaires, data analysis and interpretation.
- Dr Stephen Day (MMU), Dr Georgina Stebbings (MMU), Dr Robert Erskine (Liverpool John Moores University), Prof Craig Sale (Nottingham Trent University) Dr Philip Hennis and Prof Hugh Montgomery (both University College London): responsible for project design, collecting blood and saliva samples, conducting other measurements and administering questionnaires, data analysis and interpretation.
- Sarah Lockey, Shane Heffernan, Adam Herbert, Jon Brazier, Mark Antrobus, Peter Callus (MMU): PhD Students; responsible for collecting blood and saliva samples, conducting other measurements and administering questionnaires, data analysis and interpretation.

10) Who is funding the research?

This project is being funded by the Health, Exercise and Active Living Research Centre at Manchester Metropolitan University.

11) Who will have access to the data?

All information collected during the course of this research project will be kept confidential and will only be used for the purposes of the study. The data will be stored anonymously; only the members of the research team named above will have access to it.

The results from the study will be communicated at scientific conferences and published in peer-reviewed scientific journals some time in the future but in a manner that does not allow an individual's identity to be determined. You may obtain a copy of any publication that result from the research by contacting the Principal Investigator (see below - section 13).

12) Who do I contact if I feel my rights have been violated?

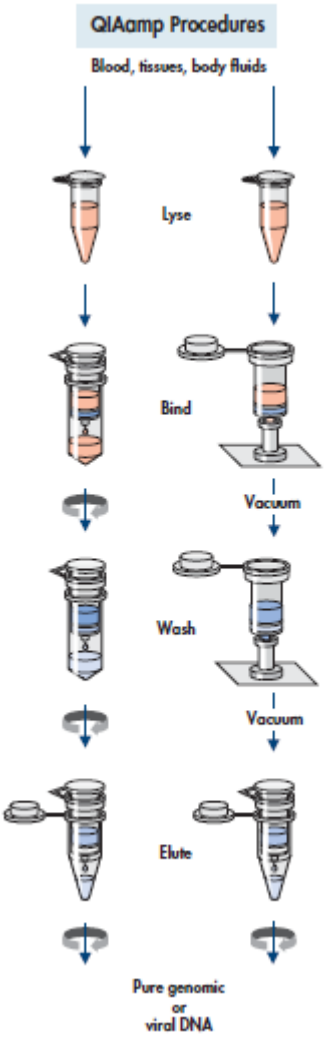
If you wish to make a complaint regarding your involvement in the study, please contact:

MMU Ethics Committee
Registrar & Clerk to the Board of Governors
Head of Governance and Secretariat Team
Manchester Metropolitan University
All Saints Building, All Saints
Manchester M15 6BH
Tel: 0161 247 1390

13) Finally, thank you for considering to participate in our research study. If you require any more information, please contact:

Dr Alun Williams
Email: a.g.williams@mmu.ac.uk
Tel: 0161 247 5523

Appendix 3: Qiacube automated protocol diagram



http://www.qiagen.com/us/resources/download.aspx?id=%7B6F5553FD-3CA5-4E2E-9DE7-43473DE3C173%7D&lang=en&ver=1

Rugby union playing positions. Adapted from *http://pinkrugby.com/wp-content/uploads/2011/01/rugby-positions.jpg*.

Position descriptions (England Rugby, 2016)

1 & 3 Prop - Along with the hooker, the loose-head and tight-head props make up what is known as the front row, which refers to their position in the scrum. To be successful, both props must be extremely strong in the neck, shoulders, upper body and legs. While stopping their side of the scrum from moving backwards, the props also support the hooker's body weight, allowing him or her to see and strike the ball when it is put into the scrum. In the lineout, props should be able to support or lift the jumper to prevent the opposition winning the ball. Away from set pieces, props help to secure the ball when a player has been tackled, so it helps if they can combine their power with a degree of mobility. You'll also often see them used as battering rams in attack, receiving short passes after a ruck or maul and hitting the opposition defence at pace in an attempt to occupy the defenders and make space for their own backs.

2 Hooker - Lining up in the scrum between the two props, the hooker is one of the forwards' key decision-makers. He or she will coordinate the timing at the scrum, and is also responsible for winning possession in the scrummage by hooking the ball back through the props' legs. To allow the hooker to do this effectively, the props support much of the hooker's weight, leaving him or her free to concentrate on hooking the ball back, rather than pushing against the opposition forwards. For this reason the hooker is often the smallest member of a front-row trio. At the lineout he or she is responsible for throwing the ball in and must be able to accurately hit the lineout jumper who is expecting the ball. In open play the hooker plays a similar role to the props, securing possession at rucks and mauls, or taking short 'crash' passes.

4 & 5 Lock - The second row forwards (also known as locks) are the power house of the scrum and the target men in the lineout, meaning that they need to be tall, powerful players with excellent scrummaging technique and pinpoint timing. In open play the second row's duties have evolved from being support players at rucks and mauls to ball carriers.

6 & 7 Flankers - Open-side and blind-side flankers are often considered to be the players with the fewest set responsibilities, but as such must be excellent all-rounders. Speed, strength, fitness, tackling and handling skills are all vital. Flankers are more often than not at the centre of the action – winning balls at the ruck and maul, collecting short passes from tackled players and making their own big tackles in open play. The open-side flanker plays on the far side of the scrum from the touchline and is often smaller in size than their blind-side partner, making them more mobile around the pitch. The blind-side flanker tends to have bigger, more physical role around the pitch, and also acts as a target jumper in the lineout.

8 Number Eight - Support play, tackling and ball-carrying are the No.8's areas of expertise, making his or her duties similar to the two flankers. Together the trio forms a unit called the back row. Binding on right at the back of the scrum, the No.8 is also the only player from the forwards who is allowed to pick the ball up from the base of the scrum. It is a move that is often used to gain vital yards when a team is scrummaging close to the opposition try line, and for it to be truly effective the No.8 must be an explosive, dynamic runner.

9 Scrum-Half - Acting as the link between the forwards and the backs, the scrum half is a key player when it comes to building attacks. Playing just behind the forwards, a good scrum half will control exactly when the ball is fed out to the backs from the rear of a scrum, ruck or maul. A scrum half needs good vision, speed and awareness, quick hands and lightning reactions. They tend to be one of the smaller players on the pitch and so rely on protection from their own forwards

10 Fly-Half - Arguably the most influential player on the pitch The fly half must orchestrate the team's back line, deciding what rehearsed moves to put into action and reacting to gaps in defence. He or she is also the main target for the defending team's open-side flanker and so must be strong in the tackle. The fly half has to be able to relieve territorial pressure by kicking down the field into touch, and is often the team's designated place kicker for conversions, penalties and drop goal attempts. In defence he or she will marshal the backs to ensure each opposition player is covered, and a strong-tackling fly half can snuff out opposition attacks before they start.

11 & 14 Wing - Playing out wide on the side of the pitch, the winger is a team's finisher in attack. A winger is also often the last line of defence when they don't have the ball and as such, pace is their major resource.

12 & 13 Centre - The inside centre – who stands closest to the fly half when the backs line up – and the outside centre tend to be strong, dynamic runners with a good eye for exposing gaps in the opposition defence. In attack they tend to run very direct lines. The centres take on their opposite number in an attempt to either break the defensive line, or draw in enough opposition defenders to create space and try-scoring opportunities for their team-mates. As such they need to be strong and powerful, and when attack turns into defence, they must also be accomplished at tackling. The inside centre is often the more creative in a centre pairing and should be able to pass and kick nearly as well as the fly half. In either defence or attack, the inside centre tends to be all action – dishing out the tackles and then drawing the opposition defence. Meanwhile, the outside centre tends to be the faster of the two and the ability to offload the ball quickly to the wingers is also vital.

15 Full-back - Lining up behind the entire back line, the full back is the closest thing that rugby has to a sweeper in defence. But they also receive deep kicks from the opposition, so they must be comfortable catching high balls and launching attacks from the resulting possession. As such, the full back must have enough tactical awareness to recognise when to counter-kick, and when to run with the ball, often from deep within his or her half.

Reference

England Rugby (2016). Positions guide, ed. Rugby Football Union. Rugby Football Union,, *http://www.englandrugby.com/*.

Single team sports

Field hockey

Just a single study reports the genetic characteristics of field hockey players. Gronek et al. (2013) investigated the association between ACE I/D and select physiological variables in elite Polish field hockey athletes (n = 47), however did not make comparison to a control group. No significant associations were identified, although, the ID genotype group (n = 20) tended towards greater estimated peak power (P = 0.085) and the II genotype group (n = 14) tended towards the greatest VO_{2max} (P = 0.081). While the sample size was extremely low and was likely the cause of non-significant results, an estimate of the effect would have been useful to truly understand the research findings (the absence of which is common within sports genomics; Figure 8). Importantly, the sample consisted of 'elite' athletes (Polish national team) which is an important criterion in studying the genetic variation among athletes.

Volleyball

The first of two investigation to consider volleyball athletes examined the effect of ACTN3 R577X variation on estimated leg power in elite players (n = 66; male and female) compared to a control group (n = 334; Ruiz et al., 2011b). There was no association identified between athletes and controls for genotypic frequency, nor a genotype-phenotype association with leg power for ether athletes or the larger cohort of controls. However, the athlete group did deviate from Hardy-Weinberg equilibrium This deviation could be for a number of reasons, however was most likely due to the small sample or genotyping error as they are the most common reasons for deviation (Sen & Burmeister, 2008). Salles et al. (2015) performed a fascinating investigating into the genetic variation of tendinopathy in elite volleyball athletes, examining 23 SNP's in the BMP4, FGF3, FGF10 and FGFR1 genes. A significant association between tendinopathy and BMP4 rs2761884 (P = 0.03) was observed with a 2.4 times greater susceptibility to tendinopathy (OR = 2.39) under presence of the T allele. Furthermore, haplotype analysis of chromosome 14 revealed an association between 'TTGGA' in the BMP4 gene (P = 0.01, OR = 1.92) and tendinopathy, with a weaker haplotype association in chromosome 11 (FGF3 'TGGTA' haplotype; P = 0.05, OR = 1.40). These results may have good future applications in the study of tendon injury and indeed tendon properties. While this study did not investigate genetic frequency variation with athlete status, high quality specific phenotype studies are important advances in team sport genomics, however significantly greater sample sizes are required.

Basketball

Two reports considering the genetic characteristics of basketball athletes exist, the first of which (Garatachea et al., 2014) investigated the association between ACTN3 R577X genotype and athlete status as well as estimated explosive leg power (jump performance) in a mixed sex group of elite Spanish basketball athletes (n = 102, 61 male) and a control sample (n = 283). Unsurprisingly, given previous data of a similar design (Ruiz et al., 2011b), their results indicated no relationship between genotype and phenotype (athlete status and explosive leg power), however the authors failed to consider positional differences. In senior male basketball athletes, physiological and anthropometric characteristics differ

considerably across playing positions. For example, 10 m sprint time can differ by ~0.26 s (13%; point guards versus centres), countermovement jump by ~12 cm (22%; small forwards versus power forwards) and squat maximum by ~32 kg (16%; point guards versus centres; Abdelkrim et al., 2010; Scanlan et al., 2014). These and other physiological differences between playing positions could be reflected in genetic characteristics, yet might not be apparent when analysing athletes as a whole cohort (Egorova et al., 2014). While investigating genetic variation of female athletes is necessary to understand the complex athletic traits in female sport, combining male and female athletes as a single cohort, in team sports such as basketball (Garatachea et al., 2014) may increase the possibility of false negative results, which would likely be due to differences in certain physiological characteristics between male and female athletes - particularly jump performance (Ziv & Lidor, 2009).

A more recent investigation of the GNB3 gene in elite Caucasian Turkish basketball players (n = 72) showed that the rs5443 CC genotype had the greatest VO2peak (CC = 60.1 versus TC = 56.7 $ml \cdot kg^{-1} \cdot min^{-1}$; P = 0.007) and isokinetic knee flexion (CC = 175 versus TT = 146 N.m; P < 0.05), with the TT genotype showing the greatest drop in anaerobic power, as determined by a 30 s Wingate test (TT = 62.9 versus TC = 54.2%; P < 0.001; Gülyaşar et al., 2014). These results suggest that the T allele is associated with a reduction in exercise performance, specifically in elite basketball (Gülyaşar et al., 2014). However, those results were not supported by other elite athletic/ancestral groups investigating rs5443 (Eynon et al., 2009; Eynon et al., 2011a; Ruiz et al., 2011a). This illustrates the importance of exploring individual sports as isolated cohorts. It is possible that the results of Gülyasar et al. (2014) may be false positives possibly due to low sample sizes, but it could also be the case that the GNB3 rs5443 C allele may be of an advantage only in basketball or similar sports heavily dependent on the above phenotypes.

Cricket

Two peer-reviewed publications have considered molecular genetic variation of cricket athletes, with both studies consisting of the same cohort (non-elite athletes n = 14 and controls n = 17; Djarova et al., 2011a; Djarova et al., 2011b). The authors attempted to associate health markers (blood pressure, C-reactive protein, uric acid, lactate and selected physiological measures) to individual genetic variants (ACE I/D, ACTN3 R577X and TNF G308A). This sample is exceptionally small for genetic analysis and this was evident by the lack of ACTN3 XX (which is present in the Zulu/Bantu population (~1%; Yang et al., 2007)) and ACE II genotypes. There was no genetic association reported between athletes and controls, but an interesting finding was that the ACE D, ACTN3 R and TNF A alleles were associated with lower C-reactive protein levels for both cricketers and controls (P = 0.001). Contrastingly, TNF has been shown to have the opposite association, in that lower C-reactive protein levels were associated with the G allele in a much larger sample (n = 456) - including non-Caucasian participants (n = 232; Lakka et al., 2006). This suggests the possibility that the low sample size studied by Djarova et al induced type 1 statistical error and as such false results.

Soccer (football)

Within team sports genomics, soccer has received the most scientific interest, with 15 peer-reviewed publications focusing exclusively on the molecular genetic characteristics of soccer athletes (Table 1 and Figure 8). Many polymorphisms have been investigated in

comparison to other athletes and controls, but also for a variety of phenotypic measures. In the first investigation to assess soccer athletes, 60 soccer players comprising of Spanish first (n = 18), second (n = 27) and third division (n = 15) competitors and a group of "world class" endurance athletes were investigated for their association with ACTN3 R577X genotype (Santiago et al., 2008). Soccer athletes had greater RR genotype frequencies than the endurance and control groups (Santiago et al., 2008), with a more recent publication supporting these findings (Pimenta et al., 2013). Furthermore, Egorova et al. (2014) observed a higher frequency of R allele carries in 57 elite players, but not in a mixed standard cohort (n = 240) of Russian athletes. This study design has been used to assess a number of other genetic variants (Juffer et al., 2009; Eynon et al., 2012; Pimenta et al., 2013; Egorova et al., 2014; Gineviciene et al., 2014) and generally supports genetic differences between soccer athletes and controls. However, most of these studies compared soccer athletes to controls as two whole cohorts and did not account for positional differences. Those that did identified significant genetic variation across playing position (Egorova et al., 2014; Gineviciene et al., 2014), which is logical because in soccer, like most other team sports, physiological demands differ considerably across playing position (Bradley et al., 2013). The consideration of positional specificity regarding team sport athlete genomics is vital to truly elucidate the genetic contribution to performance, injury risk, etc.

Two variants of the COL5A1 gene, in combination with each other (haplotype), have been linked (P = 0.048) to anterior cruciate ligament (ACL) rupture among 91 injured players compared to uninjured players (n = 143), of the same standard (Ficek et al., 2013). While the statistical significance is not large, comparing medically diagnosed ACL ruptured athletes to non-injured athletes from the same athletic level, makes this study design strong. It is worth noting the absence of a significant result when each variant was assessed separately, with the authors showing that the sample was too low to detect an association (power = 47% at 95% CI) - the inclusion of an effect estimate would have assisted interpretation of the data. Interestingly, the same sample and design was used to investigate another collagen gene variation (COL2A1 rs970547) and in contrast to their hypothesis found no associations between injured and non-injured athletes (Ficek et al., 2014). More common/less severe, non-contact soft musculoskeletal tissue injuries (NCSMTIs) have been studied in 73 Spanish athletes, including some elite players, of mixed geographic ancestry. Four genetic variants (IGF2 rs3213221, CCL2 rs2857656, COL5A1 rs12722 and ELN rs2289360) were associated with degree of injury (mild, moderate or severe) and recovery time (days) following injury, specifically muscle and ligament tissue injuries (Pruna et al., 2013). For example, depending on the genetic variant within the ELN rs2289360 (associated with ligament injury and tissue repair) the recovery time was 70% less for the beneficial genotype, although there were only 10 cases - but as preliminary data these findings are interesting (Pruna et al., 2013). Further caution must be taken when considering these results because individual genetic variants are known to differ in allele frequency between different geographical ancestry populations (1000 Genomes Consortium, 2012) which may be a factor given the combination of athletes from three geographical ancestries (European white, black African and Hispanic).

Likewise, other polymorphisms in the VDR gene have been shown to account for 19% (P = 0.041) of musculoskeletal injury severity in elite Italian soccer players (Massidda et al., 2015b). Furthermore, utilising a total genotype score (TGS) of 5 SNPs (COL5A1 rs12722,

MTC1 rs1049434, VDR Apa I, VDR Bsm I, and VDR Fok I) and in combination with player training volume, Massidda et al. (2014a) predicted 10% of the variability in injury incidence within the same sample of elite soccer athletes. Substantially larger replications of these types of study designs in soccer and other team sports may show future predictive value within sports genomics regarding injury risk.

Multiple team sports

Interestingly, three publications have concentrated specifically on combining groups of team sport athletes, in comparison to either controls (Ahmetov et al., 2013) or to other athletic groups and controls (Eynon et al., 2014; Massidda et al., 2015a). Ahmetov et al. examined a polymorphism in the peroxisome proliferator-activated receptor α (PPARA) gene (regulator of lipid and glucose metabolism) in a large mixed male and female non-elite athlete cohort (n = 655) from a variety of sports. They found that the C allele, previously associated with a higher proportion of type II muscle fibres and power-orientated athletes (Ahmetov et al., 2006), was over-represented in the whole cohort compared to controls (P = 0.0009), although when each sport was analysed individually only football (n = 241) was different from controls (P < 0.0001). These results may be due to the low sample size of the other team sport groups (the largest being basketball, n = 85) or the mix of male and female athletes (specific gender proportions of each group were not presented). Importantly the authors analysed each sport separately, in addition to the whole cohort, which is something rarely seen in the field and important for interpretation of the overall results, as just described.

Secondly, ACTN3 R577X was assessed in a combined cohort of Spanish, Polish and Russian Caucasian athletes, of which 72% were classified as elite (Eynon et al., 2014). The authors compared a group of team sport athletes (n = 205) with endurance athletes (n = 305), sprint athletes (n = 378) and a control group (n = 568). No associations were found between genotype and athletic status of team sport athletes, although the RR genotype was less common in team sport athletes then strength athletes (OR = 0.58, 95 % CI 0.34–0.39, P = 0.045). The greatest concern regarding these results was the five separate sports considered in the team sport cohort (soccer, ice hockey, field hockey, handball and water polo), each with very different physiological demands and anthropometric characteristics. While combining team sport athletes together increases sample size, analysing sports with vastly different physiological and anthropometric characteristics - which is likely to be evident at the molecular level – may dissolve important data. The same result – no association between ACTN3 and team sport athletes status - was seen in an Italian cohort (Massidda et al., 2015a), but with similar methodological limitations. While Ahmetov et al. (2013), Eynon et al. (2014) and Massidda et al. (2015a) were the first to specifically analyse the genetic variation among team sports athletes in a relatively large cohort, some methodological considerations (discussed below) need to be addressed before similar studies involving team sport athletes should be reported.

Table 1 Key characteristics of research articles included in the present chapter, separated into articles that investigated a single team sport (Single Sport Articles) or multiple sports including some team sport athletes (Multiple Sport Articles).

Single Sport Articles					
Sport(s)	Participants	Genetic variant(s)	Statistics	Relevant associations	Reference
Rugby					

Rugby union (n = 431), rugby league (n = 83)	507 athletes (all M; all elite) and 710 controls (61% M) UK South Africa	*ACE* ID, *ACTN3* R577X rs1815739	HW; Yes – all athletes and controls Sig test; $\chi 2$ MC; Benjamini-Hochberg EE; OR	No associations with *ACE* ID. For *ACTN3*, no association was identified in the whole group. The X allele was overrepresented in RU forwards compared to backs (P = 0.02) and controls (P = 0.02). In the back three, the R allele was more common than controls (P = 0.04) and forwards (P = 0.01).	(Heffernan *et al.*, 2016)
Rugby union (n = 102)	102 athletes (All M; non-elite) and 110 controls (non-rugby athletes) UK	*ACTN3* R577X rs1815739	HW; NS Sig test; $\chi 2$, t tests, ANCOVA MC; NS EE; NS	No association with genotype and physiological variables. There was a tendency for the grouped centres/wingers/full backs for greater RR genotype (P = 0.066) and R allele frequency (P = 0.059).	(Bell *et al.*, 2012c)
Rugby union (n = 109)	109 athletes (All M; non-elite) and 108 controls (non-rugby athletes) UK	*ACE I/D*	HW; Yes Sig test; $\chi 2$, t tests, MC; NS EE; NS	No difference in genotype frequency.	(Bell *et al.*, 2010)
Rugby union (n = 68)	68 athletes (all M; non-elite) UK	*ACE I/D*	HW; NS Sig test; $\chi 2$ MC; NS EE; NS	Backs with ID had greater force (Wt %) and power output ($W \cdot kg^{-1}$) than forwards (P = 0.001; P = 0.034). Backs with DD genotype had greater jump displacement (m) and velocity ($m \cdot s^{-1}$) than forwards (P = 0.049, P = 0.007).	(Bell *et al.*, 2009)
Rugby union (n = 17)	17 athletes (6 M & 11 F; non-elite) South East Asian	*ACE I/D*	HW; Yes Sig test; NS MC; NS EE; Multivariate logistic regression	The likelihood of having a high aerobic capacity was 14.3 fold (OR = 14.27, P = 0.03) greater among subjects with the II genotype compared to ID.	(Goh *et al.*, 2009)
Field Hockey					
Field hockey (n = 47)	47 athletes (all M; all elite) Polish	*ACE I/D*	HW; NS Sig test; ANCOVA MC; NS EE; NS	No differences between any physiological variables and genotype frequency.	(Gronek *et al.*, 2013)
Volleyball					
Volleyball (n = 138)	138 athletes (all M; all elite) Brazilian	*BMP4* rs2761884, rs17563, rs2855529, rs2071047, rs762642; *FGF3* rs7932320, rs1893047, rs12574452, rs4631909, rs4980700; *FGF10* rs1448037, rs900379, rs1011814, rs593307; *FGFR1* rs13317	HW; NS Sig test; $\chi 2$, t tests, Mann–Whitney test MC; Bonferroni EE; OR	*BMP4* rs2761884 was associated with tendinopathy (P = 0.03). Athletes with the risk genotype had 2.4 times more susceptibility to tendinopathy (OR = 2.39). Haplotype TTGGA in *BMP4* (P = 0.01) had greater risk.	(Salles *et al.*, 2015)

Volleyball (n = 66)	66 athletes (31 M & 35 F; all elite) and 334 controls (243 M & 91 F) Spanish	*ACTN3* R577X rs1815739	HW; Yes for controls – No for athletes Sig test; $\chi 2$, ANCOVA, ANOVA MC; NS EE; OR	No difference in leg power by genotype.	(Ruiz *et al.*, 2011b)
Basketball					
Basketball (n = 72)	72 athletes (all M; all elite) Turkish	*GNB3* rs5443	HW; Yes Sig test; ANOVA, ANCOVA MC; NS EE; NS	CC genotype showed the greatest VO_{2peak} ($P = 0.007$) and isokinetic knee flexion ($P < 0.05$), with the TT genotype showing the greatest drop in percentage anaerobic power ($P < 0.001$).	(Gülyaşar *et al.*, 2014)
Basketball (n = 102)	102 athletes (61 M, 41 F; all elite) and 283 controls (216 M & 67 F) Spanish	*ACTN3* R577X rs1815739	HW; Yes – all athlete Sig test; $\chi 2$, ANOVA MC; NS EE; OR	No association between *ACTN3* variants and explosive leg power.	(Garatachea *et al.*, 2014)
Soccer					
Soccer (n = 54)	54 athletes (all M; all elite) Italian	*VDR* (FokI, ApaI and BsmI)	HW; Yes – all athlete Sig test; ANOVA, ANCOVA, R^2 MC; NS EE; NS	ApaI allele was associated with severity of injury ($P = 0.041$).	(Massidda *et al.*, 2015b)
Soccer (n = 246)	246 athletes (all M; 21% elite, 33% sub-elite) and 872 controls (all M) Russian	*ACE I/D*, *ACTN3* R577X rs1815739, *PPARA* rs4253778, *PPARG P12A* rs1801282, *PPARGC1A*, G482S rs8192678, *PPARD*, rs2016520, *TFAM* rs1937, *UCP2* A55V rs660339	HW; Yes – all athlete groups and controls Sig test; $\chi 2$, Fisher's exact MC; Holm–Bonferroni EE; NS	*ACE* D, *ACTN3* R, *PPARA* C and *UCP2* Val alleles were associated with football player status and combined the TGS was higher in football athletes ($P < 0.0001$) than in controls. Positional differences were between midfielders, goalkeepers and controls, with TGS ($P = 0.0023$, $P = 0.0004$).	(Egorova *et al.*, 2014)
Soccer (n = 243)	243 athletes (all M) Polish	*COL12A1* rs970547	HW; Yes Sig test; $\chi 2$, Fisher's exact MC; NS EE; OR	No difference between ACL ruptured athletes and non-rupture. G allele showed a greater likelihood of ACL injury and was 0.82 times higher ($P < 0.00001$) than in the non-rupture group.	(Ficek *et al.*, 2014)
Soccer (n = 140)	140 athletes (all M; 59% elite) Brazilian	*ACTN3* R577X rs1815739	HW; Yes Sig test; $\chi 2$ MC; NS	Greater RX and lower XX in the U-14's (n = 43) and professionals (n = 83) compared to the armature athletes ($P < 0.05$).	(Coelho *et al.*, 2014)

			EE; NS		
Soccer (n = 64)	64 athletes (all M; all elite) Italian	*COL5A1* rs12722, *MTC1* rs1049434, *VDR* (Apa I, Bsm I, Fok I)	HW; Yes Sig test; ANOVA MC; NS EE; NS	The combined influence of training volume and TGS predicted injury rate, explaining 10% of the variability in injury incidence (P = 0.03).	(Massidda *et al.*, 2014a)
Soccer (n = 60)	60 athletes (all M; all sub-elite) and 30 controls Italian	*PPARA* rs4253778	HW; Yes Sig test; $\chi 2$, t test MC; NS EE; NS	Greater G allele in soccer athletes compared to controls (P = 0.04).	(Proia *et al.*, 2014)
Soccer (n = 199)	199 athletes (all M; Sub-elite) and 167 controls (all M) Lithuanian	*ACE* rs1799752 *PPARGC1A* rs8192678 *PPARA* rs4253778	HW; Yes – athletes, No – controls for ACE Sig test; $\chi 2$, Fisher's exact MC; NS EE; OR	*ACE* DD genotype frequency was lower in defenders (P = 0.033) and midfielders (P = 0.012) compared to controls. *PPARG C1A* and *PPARA* differed between forwards and controls (P = 0.044, P = 0.034). Grouped genotypes *ACE* II + *PPARA* GC were at greater odds of appearing in the players group (OR = 2.83, P = 0.047).	(Gineviciene *et al.*, 2014)
Soccer (n = 68)	68 athletes (all M; non-elite) and 100 controls (all M) Egyptian	*ACE I/D*	HW; NS Sig test; $\chi 2$, Fisher's exact MC; NS EE; NS	DD genotype had higher left ventricular ejection fraction (%; P = 0.03), right ventricular diameter in diastole (P = 0.04) and pulmonary artery systolic pressure (P = 0.02) than other genotypes for footballers compared to controls.	(Saber-Ayad *et al.*, 2013)
Soccer (n = 73)	73 athletes (all M; all elite) Spanish	*ELN* rs2289360, *TTN* rs2742327, *SOX 15* rs4227, *IGF 2* rs3213221, *CCL* rs2857656, *TNC* rs2104772, *COL1A1* rs1800012, *COL5A1* rs12722	HW; NS Sig test; $\chi 2$, Fisher's exact MC; Benjamini-Hochberg EE; NS	*IGF2*, *CCL* and *ELN* were associated with degree of injury (P = 0.034, P = 0.026, P = 0.009), and *ELN* associated with recovery time (P = 0.043).	(Pruna *et al.*, 2013)
Soccer (n = 200)	200 athletes (all M; all elite) Brazilian	*ACTN3* R577X rs1815739	HW; NS Sig test; ANOVA MC; NS EE; NS	RR genotype jumped higher and ran faster, XX had higher VO_{2max} (P < 0.05). Higher RR genotype than the normal control population (no controls analysed in this study).	(Pimenta *et al.*, 2013)
Soccer (n = 91)	91 athletes (all M; all elite) and 141 controls (also athletes) Polish	*COL1A1* rs1800012, rs1107946	HW; Yes Sig test; $\chi 2$, Fisher's exact MC; NS EE; OR	G-T haplotype was associated with lower risk of ACL rupture (P = 0.048).	(Ficek *et al.*, 2013)
Soccer (n = 37)	37 athletes (all M; all elite)	*ACTN3* R577X rs1815739	HW; NS	Higher IL-6 concentrations for RR genotype post ECC training (P < 0.05). RX and XX had higher CK	(Pimenta *et al.*, 2012)

	Brazilian		Sig test; ANOVA MC; NS EE; NS	activity than RR post ECC exercise (P < 0.05). XX had the greatest hormonal changes post ECC training.	
Soccer (n = 60), endurance athletes (n = 100), power athletes (n = 53)	213 athletes (all M; all elite) and 100 controls (all M) Spanish	*NOS2* 786T/C rs2070744	HW; Yes – controls and soccer, No – endurance and power. Sig test; $\chi 2$, MC; Bonferroni EE; logistic regression analysis	Soccer athletes had lower TT than controls, power and endurance athletes (P = 0.02).	(Eynon *et al.*, 2012)
Soccer (n =54), endurance runners (n = 52)	106 athletes (all M; all elite) and 123 controls (all M) Spanish	*ACE* I/D, *GDF-8* K153R, *AMPD1* C34T	HW; Yes - athletes and controls for AMPD1 and GDF-8. Yes – soccer players and controls. No – endurance runners for ACE Sig test; $\chi 2$ MC; NS EE; NS	Endurance runners had lower ID and higher II genotypes than controls (P = 0.026, P = 0.01). Soccer athletes had higher ID and lower II genotypes compared to endurance runners (P = 0.005, P = 0.001). *AMPD1* CT was lower in endurance runners compared to soccer and controls (P = 0.006, P = 0.014).	(Juffer *et al.*, 2009)
Soccer (n = 60), endurance athletes (n = 102)	162 athlete (all M; all elite) and 123 controls (all M) Spanish	*ACTN3* R577X rs1815739	HW; Yes – all athlete groups and controls Sig test; $\chi 2$ MC; NS EE; NS	Soccer athletes had higher RR genotype then other groups (P = 0.041).	(Santiago *et al.*, 2008)
Cricket					
Cricket (n = 14)	14 athletes (all M; non-elite) and 17 controls Zulu South African	*ACE* I/D	HW; NS Sig test; ANOVA, $\chi 2$, Fisher's exact MC; NS EE; NS	C-reactive protein and uric acid levels were lower in D allele carriers (P = 0.001). Knee flexion and extension torque was higher in D allele carriers (P < 0.03). *ACE* genotype differed between athletes and controls (P = 0.004), however there was an absence of the II genotype in athletes and controls.	(Djarova *et al.*, 2011a)
Cricket (n = 14)	14 athletes (all M; non-elite) and 17 controls Zulu South African	*ACTN3* R577X, *TNF* G308A	HW; NS Sig test; ANOVA, $\chi 2$, Fisher's exact MC; NS EE; NS	R allele was associated with C-reactive protein levels in cricketers (P = 0.0001) and controls (P = 0.014). *ACTN3* R allele was associated with BMI and FM in cricket players (P = 0.0001) and controls (P = 0.0007). TNF, A allele was associated with C-reactive protein levels (P = 0.0001). No difference between athletes and controls for *ACTN3*	(Djarova *et al.*, 2011b)

				(XX genotype was absent) or *TNF* G308A.	
Multiple Sport Articles					
Sport(s)	Participants	Genetic variant(s)	Statistics	Relevant association	Reference
Football (n = 218), cricket (n = 156), track and field (n = 67, running events (n = 62), rowing (n = 13), boxing (n = 2), tennis (n = 12), **hockey (n = 26)**, gymnastics (n = 7)	518 athletes (449 M and 69 F; all elite) North America UK	*RANK/RANKL/OPG* rs3018362, rs4355801, rs1021188, rs9594738	HW; Yes – *rs3018362, rs1021188, rs9594738.* No - *rs4355801.* Sig test; t tests MC; Bonferroni EE; OR	8.1% of the stress fracture group and 2.8% of the non-stress fracture group were homozygote for the rare allele of rs1021188. Heterozygotes and homozygous for the rare allele of rs3018362 were associated with stress fracture ($P < 0.05$). The rare allele of rs1021188 and individuals possessing at least one copy of the rare allele of rs4355801 were associated with stress fracture injury ($P < 0.05$).	(Varley *et al.*, 2015)
Endurance (n = 40), sprint/power (n = 64), **team sport athletes (n = 74) – Soccer (n = 64), Hockey (n = 10)**	178 athletes (all M; 57% elite) and 190 controls Italian	*ACTN3* R577X rs1815739	HW; Yes – all athlete groups and controls. Sig test; NS MC; NS EE; OR	Team sport athletes showed lower RR genotype compared to sprint/power group ($P = 0.044$).	(Massidda *et al.*, 2015a)
Volleyball (n = 22), swimming (n =43), **ice hockey (n = 34),** canoeing (n = 86)	185 athletes (all M; non-elite) Polish	*ACTN3* R577X	HW; Yes Sig test; MANOVA, $\chi 2$ MC; NS EE; NS	RR had greater relative power output in countermovement and spike jumps ($P < 0.05$).	(Orysiak *et al.*, 2015)
Football, basketball, athleticism, **volleyball, handball,** judo, wrestling, taekwondo, **rugby**	150 athletes (18 F & 132 M; all elite) and 150 controls (97 F & 57 M) Turkish	*ACTN3* R577X rs1815739	HW; NS Sig test; $\chi 2$ MC; NS EE; OR	Controls showed greater R allele individuals than the athletes group ($P = 0.009$).	(Yamak *et al.*, 2015)
Endurance (n = 84), sprint/power (n = 47), mixed (n = 73) - wrestlers, tennis players**, handball players, footballers**	204 athletes (160 M & 44 F; all elite) and 260 controls Polish	*AMPD1* rs17602729	HW; Yes – all athlete groups and controls. Sig test; t test, ANOVA, $\chi 2$, Fisher's exact MC; Bonferroni EE; NS	Sprint/power athletes had the greatest CC genotype frequency compared to endurance-orientated, mixed and control group ($P < 0.05$).	(Ginevičienė *et al.*, 2014)
Endurance (n = 688), **mixed -** Badminton players (n = 24), **basketball players (n = 109), baseball players (n = 38**), boxers (n = 143) **handball players (n = 92), ice hockey**	2664 (2262 Russians and 402 Polish; 1540 M and 1124 F) and 917 controls	*SOD2* Ala16Val rs4880	HW; Yes – all athlete groups and controls Sig test; ANOVA $\chi 2$ Spearman's correlations	*SOD2* T allele had increased activity of CK ($F, P = 0.0144$) and creatinine level ($F, P = 0.0276$; M, $P = 0.0135$). TT genotype was lower in power/strength athletes compared to controls ($P = 0.0076$) and athletes involved in low-intensity sports ($P = 0.0001$).	(Ahmetov *et al.*, 2014b)

players (n = 111), karate athletes (n = 22), taekwondo athletes (n = 18), **field hockey players (n = 19),** synchronized swimmers (n = 27), fencers (n = 64), freestyle skiers (n = 11), figure skaters (n = 76), archers (n = 24), Nordic combined athletes (n = 10), snowboarders (n = 33), **football players (n = 36)**, pentathletes (n = 23), **softball players (n = 31), rugby players (n = 48),** table tennis players (n = 11), **volleyball players (n = 115), mini-football players (n = 9), water polo players (n = 59),** white water slalomists (n = 5), wrestlers (n = 294), power (n = 321), strength (n = 203)	(558 M and 359 F) Russian Polish		MC; Bonferroni EE; NS		
Volleyball (n = 61), baseball (n = 21), alpine skiing (n = 18), speed skating (n = 19), figure skating (n = 32), kayak (n = 25), cross-country skiing (n = 35)	209 athlete (119 M & 90 F; all elite) Russian	*ACTN3* R577X rs1815739	HW; Yes Sig test; ANOVA, Spearman's correlations MC; NS EE; NS	Testosterone levels were higher in both M and F athletes with the *ACTN3* R allele than XX homozygotes (P = 0.0071 for M; P = 00.0167 for F).	(Ahmetov *et al.*, 2014a)
Endurance (n = 254), sprint power group (n = 338) – **ice hockey (n = 27)**	592 athletes (485 M & 107 F; all elite) Russian Polish	*EPAS1* rs895436, rs11689011, rs1867782, rs1867785, rs4035887	HW; Yes – athlete groups and controls. Sig test; Multivariate Adaptive Regression Splines MC; NS EE; OR	rs1867785 AA genotype was underrepresented in sprint/power athletic status (P = 0.00022). TT genotype of rs11689011 was underrepresented in sprint/power athletes. The combinations of the AA genotype in rs4035887 with either the AG or GG genotypes in rs1867785, or with the CT or CC genotypes in rs11689011, were underrepresented in two cohorts of sprint/power athletes (P < 0.005).	(Voisin *et al.*, 2014)
Endurance (n = 142), power (n = 91), mixed (n = 90) – fencing (n = 8), **soccer (n = 53)**, table tennis (n = 4), **volleyball (n = 7)**, tennis (n = 4), tae Kwando (n = 9)	323 athletes (242 M & 91 F; 76% elite) Russian	*MCT1* rs1049434	HW; Yes – controls, no – all athletes Sig test; t tests, $\chi 2$, Fisher's exact MC; Bonferroni EE; OR	A allele was higher in athletes than in controls (P < 0.0001). A allele and AA genotype were higher in rowers than in controls (P < 0.0001). Mean lactate concentration were higher in M rowers with the T allele compared with AA homozygotes (P = 0.005).	(Fedotovskaya *et al.*, 2014)

Touch football (n = 37), softball (n = 28), basketball (n = 25), badminton (n = 9)	99 athletes (all F; non-elite) Japanese	*ACTN3, UCP1* rs1800592, *UCP2, UCP3*	HW; NS Sig test; $\chi 2$ MC; NA EE; OR	*ACTN3* XX athletes that did not experience sports injuries had greater frequency than RR (P = 0.0133). R allele frequency was higher than X frequency in athletes who experienced muscle injuries (P = 0.0015), with an odds ratio of 2.52.	(Iwao-Koizumi *et al.*, 2014)
Endurance cohort (n = 305; Spanish, Polish & Russian), sprint/ power cohort (n = 378; Spanish, Polish & Russian). **Team sport cohort - Soccer players (n = 50; Spanish, n = 3; Polish), ice hockey players (n = 25; Polish, n = 59; Russian), field hockey (n = 9; Polish), handball players (n = 21; Polish, n = 36; Russian), water polo (n = 2; Russian)**	205 athletes (all M; all elite) and 568 controls Spanish Polish Russia	*ACTN3* R577X rs1815739	HW; Yes – all athlete groups and controls Sig test; NS MC; NS EE; OR	Higher RR genotype in power/sprint athletes than team sport athletes (P = 0.045). No differences between team sport athletes and controls (P = 0.765).	(Eynon *et al.*, 2014)
Endurance cohort (n = 114). Power group (n = 116), **Mixed cohort - Soccer (n = 44)**, wrestlers (n = 26), boxers (n =23), judokas (n = 19), fencers (n = 18)	360 athletes (273 M & 87 F; 168 elite, 192 sub-elite) and 191 controls. Polish	*NOS3 G894T rs1799983*	HW; Yes – all athlete groups and controls Sig test; $\chi 2$ MC; false discovery rate EE; OR	GG genotype (P = 0.0006) and G allele (P = 0.0002) were overrepresented in all athlete groups, compared to cintrols.	(Eider *et al.*, 2014)
Badminton, **basketball**, canoe, karate, wrestling, lacrosse	253 athletes (144 M & 109 F; non-elite) Japanese	*ACTN3* R577X rs1815739	HW; Yes Sig test; t tests, $\chi 2$, Fisher's exact MC; NS EE; NS	Relative peak power output was higher in the R allele group than in the XX group in M (P = 0.045), but not F athletes.	(Kikuchi *et al.*, 2014a)
Endurance (n = 126), strength/endurance (n = 161), sprint/strength (n = 640) – **football (n = 82), ice hockey (n = 70)**, strength (n = 197)	1124 athletes (757 M & 367 F; 41% elite) and 1191 controls (684 M & 507 F) Polish Russian	*BDKRB2* -9/+9	HW; Yes – athlete groups and controls Sig test; $\chi 2$ MC; Bonferroni EE; NS	No differences between athletes and controls.	(Sawczuk *et al.*, 2013)
Mixed cohort - badminton (n = 16), **baseball (n = 28)**, **basketball (n = 85), beach volleyball (n = 63)**, court tennis (n = 33), **football (n = 241)**,	665 athletes (M & F; sub-elite) and 1706 controls Russian	*PPARA* GC rs4253778	HW; Yes – controls, no – athletes Sig test; $\chi 2$ MC; Bonferroni	C allele was higher in athletes compared to controls (P = 0.0009). Football (P < 0.0001) and softball (P = 0.047) athletes, independently, had higher frequencies of C allele compared to controls.	(Ahmetov *et al.*, 2013)

futsal (n = 9), **handball (n = 24)**, **ice hockey (n = 55)**, **rugby (n = 48), softball (n = 31)**, table tennis (n = 14), **water polo (n = 18)**			EE; NS		
Speed/strength group, endurance/speed/ strength group – **field hockey**, tennis, **rugby, football, volleyball, basketball, handball**, boxing, and kickboxing, canoeing, rowing. Endurance group	156 athletes (119 M & 37 F; all elite) and 83 controls (35 M & 48 F) Polish	*UCP2* ID UTR, *UCP3* CT	HW; Yes – all athlete groups and controls Sig test; ANOVA, t tests MC; NS EE; OR	No significant differences for genetic markers and phenotype measures.	(Holdys *et al.*, 2013)
Ball game cohort - Soccer (n = 16), baseball (n = 8), basketball (n = 10), volleyball (n = 8), ice hockey (n = 8)	50 athletes (non–elite) and 33 controls Korean	*ACE* G2350A rs4343	HW; Yes – all athlete groups and controls Sig test; $\chi 2$, ANOVA MC; NS EE; NS	No significant differences for genetic markers and cardiovascular function.	(Jang & Kim, 2012)
Gymnastics (n = 17), 100-400 m running (n = 12), **soccer (n = 30)**	59 elite athletes and 31 controls Italian	*ACE* I/D, *ACTN3* R577X	HW; NS Sig test; $\chi 2$ MC; NS EE; NS	*ACE*, no significant difference. The D allele was high in all cohort (DD; D: G = 53%; 70%, R = 50%; 75%, S = 60%; 73%, C = 45%; 66%). *ACTN3*, there was an absence of the XX genotype in the gymnastic group and excess RR (P = 0.03). No difference for ether *ACE* or *ACTN3* in soccer group.	(Massidda *et al.*, 2012)
Swimming (n = 44 M: n = 25), **volleyball (n = 16 M), handball (n = 29 F)**, long distance runners (n = 42 M: n = 19 F)	125 athletes (102 M & 73 F; all elite) and 169 controls (88 M & 83F) Greek	*ACE* I/D rs1799752, *LEP* G–2548A rs7799039, *AGTR1* A1166C UTR rs12721276, *BDKRB2* rs72348790	HW; NS Sig test; Fisher's exact test MC; NS EE; NS	Higher DD in F athletes than F controls (P = 0.034), with ID higher in all controls (P = 0.027). Higher I allele in F athletes (P = 0.034). F athlete had greater DD genotype and the *BDKRB2* +9/−9 and *LEP* GA was more prevalent (P = 0.001, P = 0.021) then controls. *BDKRB2* +9/+9 genotype in F (P = 0.042) was greater than controls. Allele combinations of IG+9A and IG−9A (P = 0.017, P = 0.004) were significant compared to controls.	(Sgourou *et al.*, 2012)
Power group, endurance group – 800 m running, 400 m swimming, **hockey**, Mixed group - **Basketball,** tennis, **volleyball**	147 athletes (106 M & 41 F; 69 elite) and 131 controls Indian	*ACE* I/D	HW; NS Sig test; $\chi 2$ MC; NS EE; NS	*ACE* I allele was higher in all athletes compared to controls (P = 0.05). No difference between power and endurance athletes.	(Kothari *et al.*, 2012)
Endurance group (n = 77), power group (n = 51), mixed group (n = 65) – tennis (n = 3), **handball (n = 14),** boxing (n = 6),	193 athletes (152 M &41 F; 43 elite; 52 sub-elite) and 250 (167 M &	*ACE I/D* rs1799752, *ACTN3* rs1815739, *PPARGC1A* rs8192678, *PPARA* rs4253778	HW; *ACE* I/D – yes athletes, no controls. *ACTN3* – yes controls no athletes, *PPARGC1A*	*ACE* II genotype individuals had greater short-term peak power (W) compared to DD in endurance and power group (P = 0.026). No difference in *ACE* frequency among athletes and controls. For *ACTN3*, elite athletes (Greater R allele) differed in genotype	(Ginevičienė *et al.*, 2011)

wrestling (n = 10), **soccer (n = 32)**	63 F) controls Lithuanian		& *PPARA* – yes form both. Sig test; $\chi 2$, ANOVA, ANCOVA MC; NS EE; NS	frequency from sub-elite (P = 0.04). *PPARA* C allele was more frequent in athletes than controls (P = 0.046). No difference between sporting disciplines.	
Apparatus gymnastics (n = 13), sprint (n = 9), speed skating (n = 8), weight lifting (n = 27), throwing (n = 20), badminton (n = 7), table tennis(n = 8), taekwondo (n = 11), **field hockey (n = 33)**, **handball (n = 15)**	151 elite athletes (88 M & 63 F) and 183 controls (95 M & 88 F) Korean	*ACE* I/D	HW; Yes – all athlete groups and controls Sig test; Fisher's exact test, ANOVA MC; NS EE; OR	No significant differences.	(Kim *et al.*, 2011)
Sprint and strength athletes (n = 35), endurance athletes (n = 50), mixed group (n = 71) – **field hockey**, tennis, **rugby**, **soccer**, **volleyball**, **basketball**, **handball**, boxing, kickboxing, canoeing	156 non-elite (119 M & 37 F) and 83 controls (35 M & 48 F) Polish	*ACE* I/D	HW; Yes – all athlete groups and controls Sig test; ANOVA, t tests MC; NS EE; NS	No significant differences for genetic markers or VO_{2max}.	(Holdys *et al.*, 2011)
Long distance running (n = 21), middle distance running (n = 29), sprinting (n = 12), triathlon (n = 19), race walking (n = 7), badminton (n = 3), orienteering (n = 7), **indoor soccer (n = 3), field hockey (n = 16)**, claiming (n = 4), fencing (n = 22), rhythmic gymnastics (n = 11), golf (n = 2), **goalball (n = 3)**, riflery (n = 4), swimming (n = 16), **waterpolo (n = 2),** field events (n = 15), track events (n = 5), figure skating (n = 1), archery (n = 4), athletics (n = 7), boxing (n = 7), cycling (n = 14), canoeing (n = 4), wrestling (n = 6), artistic gymnastics (n = 12), weight lifting (n = 2), judo (n = 30), karate (n = 13), taekwondo (n = 1)	299 elite athletes (193 M & 106 F) Spanish	*ACE* I/D rs4646994	HW; No whole group Sig test; $\chi 2$, ANOVA MC; Bonferroni EE; NS	No differences in genetic markers for any sporting disciplines and cardiovascular function. DD was higher in power sports, ID was higher in endurance group (P = 0.049).	(Boraita *et al.*, 2010)

Endurance group (n = 681), strength/ power group (n = 372), mixed group (n = 484) – all-round skating (n = 68), alpine skiing (n = 13), artistic gymnastics (n = 54), **basketball (n = 33)**, boxing (n = 36), diving (n = 9), **ice hockey (n = 16)**, mountain biking (n = 10), modern pentathlon (n = 19), shooting (n = 44), ski jumping (n = 14), **soccer (n = 42)**, tennis (n = 29), wrestling (n = 97)	1537 athletes (1085 M & 452 F) and 1113 (526 M & 587 F) Russian	*TFAM* (Ser/Thr)	HW; Yes – all athlete groups and controls Sig test; $\chi 2$, Fisher's exact MC; NS EE; NS	*TFAM* 12Thr allele frequency was higher in all athletes than control (P = 0.0015). In M masters athletes, W_{max} was higher in the 12Thr allele carriers as compared to Ser/Ser (P = 0.01). No differences for any team sport athlete groups.	(Ahmetov *et al.*, 2010)
Endurance group (n = 64), speed power (n = 47) and mixed (n = 33), **team (n = 49)** – tennis (n = 3), **soccer (n = 32)**, **handball (n = 14)**	193 elite athletes (153 M & 41 F) and 250 (167 M & 83 F) controls Lithuanian	*ACE I/D* rs1799752, *ACTN3* rs1815739, *PPARGC1A* rs8192678, *PPARA* rs4253778	HW; *ACE* I/D – yes athletes, no controls. *ACTN3* – yes controls no athletes. *PPARGC1A* & *PPARA* – yes form both. Sig test; $\chi 2$, ANOVA MC; NS EE; OR	*ACE* DD - higher frequency in endurance-orientated athletes compared to speed/power-orientated. ACTN3 - M athletes showed greater R allele than M controls (P = 0.03). No difference for *PPARA* or *PPARG C1A*. M athletes with ACE II genotype had higher MM and AAMP compared to the DD athletes. Power orientated *ACE* DD genotype athletes had significantly higher AAMP than ACE II athletes than endurance and mixed groups. M athletes with *PPARA* CC, *PPARG C1A* Gly482Gly and *ACTN3* RR were associated with increased SMCP. M athletes with PPARA CC and *PPARG C1A* Gly482Ser had the greatest MM. *PPARG C1A* Gly482Gly was more frequent in endurance and mixed Athletes than others (P = 0.049).	(Ginevičienė *et al*., 2010)
Fencing (n = 5), endurance (n = 5), **basketball (n = 5)**	15 athletes – non-elite Egyptian	*ACE* I/D	HW; NS Sig test; Kruskal Wallis Test, Spearman's correlation MC; NS EE; NS	Increased *ACE* protein level in DD (fencing and basketball) and elevated cardiac parameters compared to endurance athletes (P < 0.05). *ACE* II and ID were more frequent in endurance athletes.	(Heshmat *et al.*, 2010)
Endurance group (n = 71), strength/ speed group (n = 59) and **team sport group (n = 431) –** tennis (n = 3), **handball (n = 14), field hockey (n = 21), football (n = 393)**	561 athletes (all elite) and 174 controls Lithuanian	*ACE* I/D	HW; *ACE* I/D – yes athletes, no controls Sig test; $\chi 2$ MC; NS EE; NS	Lower D allele in athletes compared to controls (P = 0.004). No difference between team sport athletes and controls (P = 0.24). No difference between football playing positions (P > 0.03).	(Ginevičienė *et al*., 2009)
Endurance group (n = 694), power group (n = 481), **mixed group (n = 248) – basketball (n = 33),**	1423 (998 M & 425 F; 235 elite) athletes and 1132	*PPARA* rs4253778, *PPARD* rs2016520, *PPARGC1A* rs8192678, PPP3R1 promoter *5I/5D*,	HW; Yes – all athlete groups and controls.	*UCP2* 55Val, *UCP3* T, *VEGFA* rs2010963 C, *NFATC4* Gly160, *PPARGC1B* 203Pro and *TFAM* 12Thr were all overrepresented in endurance group (P < 0.05).	(Ahmetov *et al.*, 2009a)

boxing (n = 30), **ice hockey (n = 17), soccer (n = 42)**, tennis (n = 29), wrestling (n = 96)	controls (537 M & 595 F) Russian	*UCP2* rs660339, *UCP3* rs1800849, *VEGFA* rs2010963, *ACE* Alu I/D, *AMPD1* rs17602729, *HIF1A* rs11549465, *NFATC4* rs2229309, *PPARG* rs1801282, *PPARGC1B* rs7732671, *TFAM* rs1937 *and VEGFA* rs699947	Sig test; $\chi 2$, Spearman's correlation, t tests MC; NS EE; NS		
Endurance group (n = 182), sprint group (n = 48), strength group (n = 69) and mixed group (n = 172) – **basketball (n = 28),** boxing (n = 15), tennis (n = 15), **ice hockey (n = 13), soccer (n = 5),** swimming 200-400m (n = 5), wrestling (n = 36)	471 (323 M & 148 F; 52 elite) athletes and 602 controls (202 M & 401 F) Russian	*VEGFR2* His472Gln rs1870377	HW; Yes – all athlete groups and controls Sig test; $\chi 2$, t tests MC; NS EE; NS	Athletes had a higher frequency of 472Gln then controls (P = 0.0032) Endurance and sprint groups had higher frequency then controls (P = 0.0006, P = 0.007). The Gln allele showed a higher proportion of slow twitch muscle fibres in controls and athletes (P = 0.037, P = 0.01).	(Ahmetov *et al.*, 2009b)
Power-orientated athletes (n = 486) – alpine skiing (n = 29), artistic gymnastics (n = 44), bodybuilding (n = 23), figure skating (n = 10), **ice hockey (n = 34),** jumping events (n = 8), powerlifting (n = 9), running 100-400m (n = 70), ski jumping (n = 18), **soccer (n = 4),** speed skating (n = 19), swimming 50-100m (n = 10), throwing events (n = 15), **volleyball (n = 9),** weightlifting (n = 55), wrestling (n = 58)	486 athletes (363 M & 123 F; 100 elite) and 1197 controls (524 M & 673 F) Russian	*ACTN3* R577X	HW; Yes – all athlete groups and controls Sig test; $\chi 2$ MC; NS EE; NS	The X allele and XX genotype was lower in athletes compared to controls (P = 0.0004, P = 0.001).	(Druzhevskaya *et al*., 2008)
Endurance group(s) (n = 609), sprint/power group(s) (n = 258), mixed (n = 289) – **basketball (n = 20),** boxing (n = 22), wrestling (n = 82), **volleyball (n = 6),** speed skating (n = 62), mountain biking (n = 10), table tennis (n = 4), pentathlon (n = 17), shooting (n = 24), tennis (n = 15), **soccer (n = 10),** fencing (n = 5), **ice hockey (n = 12)**	1256 athletes (883 M & 373 F; 386 elite) and 610 controls (387 M & 223 F) Russian	*PPARD* 294 C/T rs4253778	HW; Yes – all athlete groups and controls Sig test; $\chi 2$, fisher's exact MC; NS EE; NS	The C allele was higher in the athlete group (P = 0.0001).	(Ahmetov *et al.*, 2007)
Basketball (n = 15), soccer (n = 41), baseball (n = 31),	139 athletes (all M) – all elite and	*ACE* I/D	HW; NS	No difference in any physiological measures. No difference between athletes and controls.	(Oh, 2007)

gymnastics (n = 12), **volleyball (n = 7), ice hockey (n = 17),** judo (n = 8), marathon (n = 8)	163 controls (all M) Korean		Sig test; $\chi 2$, ANOVA MC; NS EE; NS		
Endurance group (n = 491), power group (n = 180), Mixed group (n = 155) – boxing (n = 22), **ice hockey (n = 15),** wrestling (n = 63), tennis (n = 15)	786 athletes (571 M & 215 F; 61 elite) and 1242 controls (559 M & 683 F) Russian	*PPARA* G/C rs4253778	HW; Yes – all athlete groups and controls Sig test; $\chi 2$, Spearman's correlation MC; NS EE; NS	An increasing linear trend of C allele with increasing anaerobic component of physical performance (P = 0.029). Ice hockey athletes had greater CC frequency compared to controls (P = 0.032).	(Ahmetov *et al.*, 2006)
Dancers (n = 85), endurance athletes (n = 32 F & 4 M), **ball game sports (basketball & volleyball; n = 24 F & 15 M)**, sailing & fencing (n = 8 F & 1 M), martial arts (n = 5 F & 2 M)	91 athletes – non-elite and 872 controls Israeli	*SLC6A4:* (promoter region HTTLPR and intron 2 VNTR), Arginine vasopressin receptor 1a (AVPR1a: promoter microsatellites RS1 and RS3)	HW; Yes – all athlete groups and controls Sig test; NS MC; NS EE; likelihood ratio	*AVPR1a* haplotype frequencies (RS1 and RS3) and SLC6A4 (HTTLPR and VNTR), were observed between dancers and athletes (P = 0.000044).	(Bachner-Melman *et al.*, 2005)
Middle distance runners (n = 17), **basketball (n = 10), handball (n = 18), soccer (n = 35)**	80 athletes –non-elite and 80 controls (39 M & 41 F) Turkish	*ACE* I/D	HW; NS Sig test; $\chi 2$ MC; NS EE; NS	Athletes show greater I allele frequency than controls (P = 0.026).	(Turgut *et al.*, 2004)
Basketball (n = 15), soccer (n = 25), baseball (n = 32), gymnastics (n = 12), **volleyball (n = 18),** runner (n = 4), judo (n = 9), marathon (n = 5)	120 athletes (all M) Korean	*Haptoglobin* (Hp)	HW; Yes Sig test; $\chi 2$, MC; Scheffe's multiple comparison EE; NS	Hp1-1 homozygotes had higher VO_{2max} than Hp2-2 (P = 0.008).	(Kang *et al.*, 2003)
Swimmers (n = 48), weight bearing sports (n = 84) – **volleyball, basketball, handball,** high jump	132 athletes (all M) and 80 controls Japanese	*VDR* (RFLP)	HW; NS Sig test; ANOVA MC; NS EE; NS	Weight bearing athletes with the FF genotype had greater bone mineral density then matched controls (P < 0.01).	(Nakamura *et al.*, 2002b)
Weight bearing sports (n = 44) – **volleyball, basketball, handball,** high jump, triple jump	44 athletes (all M) and 44 controls Japanese	*VDR* (RFLP)	HW; NS Sig test; Students t test, ANOVA MC; NS EE; NS	FF genotype had greater bone mineral density compared to controls (P < 0.01).	(Nakamura *et al.*, 2002a)
Hockey (n = 26), cycling (n = 25), skiing (n = 21), track and field (n = 15), swimming (n = 13),	120 athletes (81 M and 39 F) and 685 controls	*ACE* I/D	HW; NS Sig test; NS MC; NS	No difference between groups.	(Taylor *et al.*, 1999)

rowing (n = 7), gymnastics (n = 5), other (n = 8)	(347 M and 338 F) Australian		EE; OR		

If information is absent from cells within the table, this is because it was not evident from the article. **Abbreviations**; M, male; F, female; HW, Hardy-Weinberg equilibrium; Sig test, significance test; MC, multiple test correction; EE, effect estimate; OR, odds ratio; NS, not stated; AAMP, anaerobic alactic muscular power; SMCP, single muscular contraction power; MM, muscle mass; ECC, eccentric; CK, creatine kinase; BMI, body mass index; FM, fat mass; ACL, anterior cruciate ligament.
.

Table 2UK and SA participant genotype frequencies.

Gene variant	Genotype frequency	
	UK participants	**SA participants**
ACE I/D – II,ID,DD	21.0,49.5, 29.5	22.1, 47.6, 30.3
ACTN3 – RR,RX,XX	19.0, 47.5, 33.5	18.5, 50.1, 31.4
FTO – AA,AT,TT	16.4, 49.5, 34.1	17.4, 47.8, 34.8
APOE ε4 - +ε4, -ε4	29.2, 70.8	31.1, 68.9
COL5A1		
rs12722 – TT,TC,CC	27.0, 54.3, 18.7	26.4, 55.3, 18.3
rs3196378 – AA,AC,CC	27.3, 50.2, 22.5	27.1, 49.6, 23.3

Chi^2 showed no difference in genotypic frequency between UK and SA participants ($P > 0.05$).

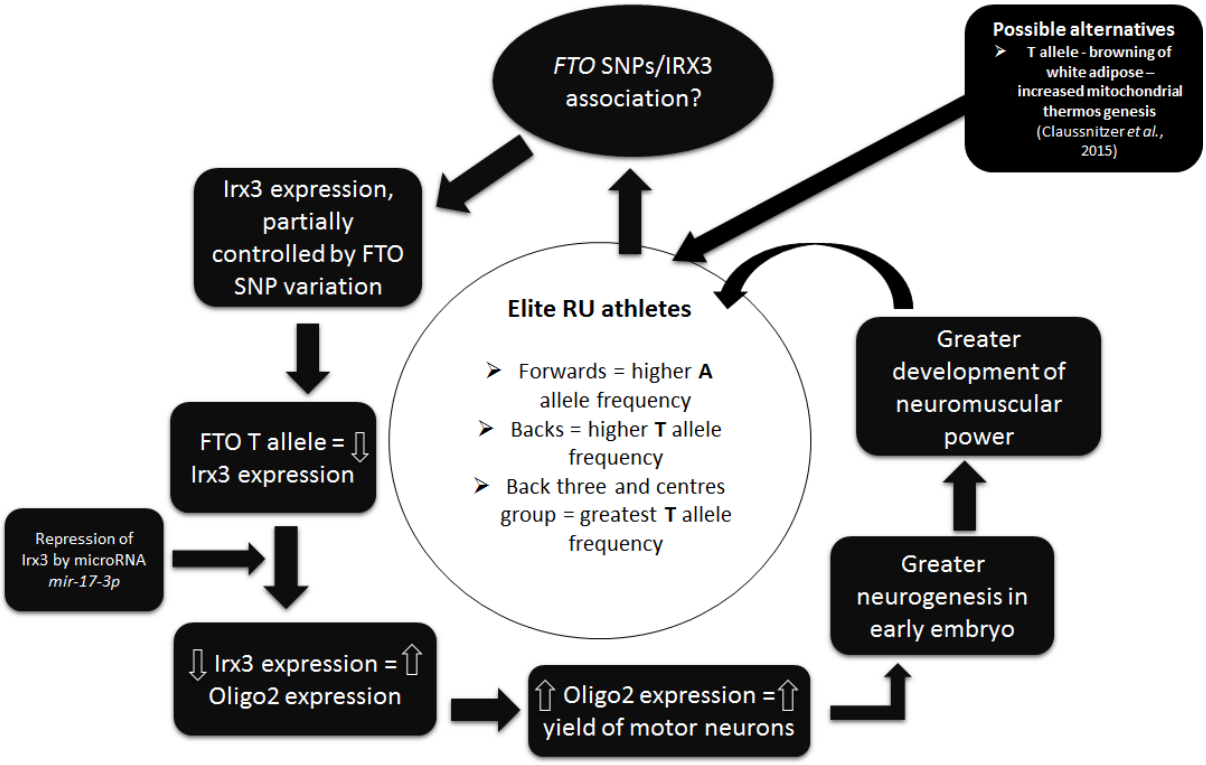


Figure 1 FTO SNP/IRX3 plausible pathway

Thesis references

References

1000 Genomes Consortium (2012). An integrated map of genetic variation from 1,092 human genomes. *Nature,* 491*(7422)*, 56-65.

Abdelkrim NB, Chaouachi A, Chamari K, Chtara M & Castagna C (2010). Positional role and competitive-level differences in elite-level men's basketball players. *The Journal of Strength & Conditioning Research* 24*(5)*, 1346-1355.

Abrahams Y, Laguette MJ, Prince S & Collins M (2013). Polymorphisms within the COL5A1 3′-UTR That Alters mRNA Structure and the MIR608 Gene are Associated with Achilles Tendinopathy. *Annals of Human Genetics,* 77*(3)*, 204-214.

Ahmetov, II, Kulemin NA, Popov DV, Naumov VA, Akimov EB, Bravy YR, Egorova ES, Galeeva AA, Generozov EV, Kostryukova ES, Larin AK, Mustafina LJ, Ospanova EA, Pavlenko AV, Starnes LM, Żmijewski P, Alexeev DG, Vinogradova OL & Govorun VM (2015). Genome-wide association study identifies three novel genetic markers associated with elite endurance performance. *Biology of Sport,* 32*(1)*, 3-9.

Ahmetov, II, Popov DV, Missina SS, Vinogradova OL & Rogozkin VA (2010). Association of the mitochondrial transcription factor (TFAM) gene polymorphism with physical performance of athletes. *Fiziologiia Cheloveka's* 36*(2)*, 121-125.

Ahmetov, II, Williams AG, Popov DV, Lyubaeva EV, Hakimullina AM, Fedotovskaya ON, Mozhayskaya IA, Vinogradova OL, Astratenkova IV, Montgomery HE & Rogozkin VA (2009a). The combined impact of metabolic gene polymorphisms on elite endurance athlete status and related phenotypes. *Human Genetics,* 126*(6)*, 751-761.

Ahmetov I, Astratenkova I & Rogozkin V (2007). Association of a PPARD polymorphism with human physical performance. *Molecular Biology,* 41*(5)*, 776-780.

Ahmetov I, Donnikov A & Trofimov D (2014a). ACTN3 genotype is associated with testosterone levels of athletes. *Biology of Sport,* 31*(2)*, 105-108.

Ahmetov I, Naumov V, Donnikov A, Maciejewska-Karłowska A, Kostryukova E, Larin A, Maykova E, Alexeev D, Fedotovskaya O & Generozov E (2014b). SOD2 gene polymorphism and muscle damage markers in elite athletes. *Free Radical Research,* 48*(8)*, 948-955.

Ahmetov II, Druzhevskaya AM, Lyubaeva EV, Popov DV, Vinogradova OL & Williams AG (2011). The dependence of preferred competitive racing distance on muscle fibre type composition and ACTN3 genotype in speed skaters. *Experimental Physiology,* 96*(12)*, 1302-1310.

Ahmetov II, Egorova E, Gabdrakhmanova L & Fedotovskaya O (2016). Genes and Athletic Performance: An Update. In *Genetics and Sports*, 2nd edn, ed. Posthumus M & Coollins M, pp. 41-54. Karger Publishers.

Ahmetov II, Egorova ES & Mustafina LJ (2013). The PPARA gene polymorphism in team sports athletes. *Central European Journal of Sport Sciences and Medicine,* 1*(1)*, 19-24.

Ahmetov II & Fedotovskaya ON (2012). Sports genomics: Current state of knowledge and future directions. *Cellular and Molecular Exercise Physiology,* 1*(1)*, e1.

Ahmetov II, Hakimullina AM, Popov DV, Lyubaeva EV, Missina SS, Vinogradova OL, Williams AG & Rogozkin VA (2009b). Association of the VEGFR2 gene His472Gln polymorphism with endurance-related phenotypes. *European Journal of Applied Physiology,* 107*(1)*, 95-103.

Ahmetov II, Mozhayskaya IA, Flavell DM, Astratenkova IV, Komkova AI, Lyubaeva EV, Tarakin PP, Shenkman BS, Vdovina AB & Netreba AI (2006). PPARα gene variation and physical performance in Russian athletes. *European Journal of Applied Physiology,* 97*(1)*, 103-108.

Alfred T, Ben-Shlomo Y, Cooper R, Hardy R, Cooper C, Deary IJ, Gunnell D, Harris SE, Kumari M & Martin RM (2011). ACTN3 genotype, athletic status, and life course physical capability: meta-analysis of the published literature and findings from nine studies. *Human Mutation,* 32*(9)*, 1008-1018.

Almeida S, Barros C, Moraes M, Russo F, Haro A, Rosa T, Alves M, Pesquero J, Carmona A & Bacurau R (2010). Plasma Kallikrein and Angiotensin I-converting enzyme N-and C-terminal domain activities are modulated by the insertion/deletion polymorphism. *Neuropeptides,* 44*(2)*, 139-143.

Altinisik J, Meric G, Erduran M, Ates O, Ulusal AE & Akseki D (2015). The BstUI and DpnII variants of the COL5A1 Gene are associated with Tennis Elbow. *The American Journal of Sports Medicine,* 43*(7)*, 1784-1789.

Alvarez R, Terrados N, Ortolano R, Iglesias-Cubero G, Reguero JR, Batalla A, Cortina A, Fernández-García B, Rodríguez C & Braga S (2000). Genetic variation in the renin-angiotensin system and athletic performance. *European Journal of Applied Physiology,* 82*(1)*, 117-120.

Amir O, Amir R, Yamin C, Attias E, Eynon N, Sagiv M, Sagiv M & Meckel Y (2007). The ACE deletion allele is associated with Israeli elite endurance athletes. *Experimental Physiology,* 92*(5)*, 881-886.

Andreasen CH, Stender-Petersen KL, Mogensen MS, Torekov SS, Wegner L, Andersen G, Nielsen AL, Albrechtsen A, Borch-Johnsen K & Rasmussen SS (2008). Low physical activity accentuates the effect of the FTO rs9939609 polymorphism on body fat accumulation. *Diabetes,* 57*(1)*, 95-101.

Appleby B, Newton RU & Cormie P (2012). Changes in strength over a 2-year period in professional rugby union players. *The Journal of Strength & Conditioning Research,* 26*(9)*, 2538-2546.

Austin D, Gabbett T & Jenkins D (2011a). The physical demands of Super 14 rugby union. *Journal of Science and Medicine in Sport,* 14*(3)*, 259-263.

Austin D, Gabbett T & Jenkins D (2011b). Repeated high-intensity exercise in professional rugby union. *Journal of Sports Sciences,* 29*(10)*, 1105-1112.

Austin SR, Dialsingh I & Altman N (2014). Multiple Hypothesis Testing: A review. *J Indian Soc Agricultural Stat,* 68, 303-314.

Bachner-Melman R, Dina C, Zohar AH, Constantini N, Lerer E, Hoch S, Sella S, Nemanov L, Gritsenko I & Lichtenberg P (2005). AVPR1a and SLC6A4 gene polymorphisms are associated with creative dance performance. *PLoS Genetics,* 1*(3)*, e42.

Barr MJ, Sheppard JM, Gabbett TJ & Newton RU (2014). Long-term training-induced changes in sprinting speed and sprint momentum in elite rugby union players. *The Journal of Strength & Conditioning Research,* 28*(10)*, 2724-2731.

Beggs AH, Byers TJ, Knoll J, Boyce FM, Bruns G & Kunkel LM (1992). Cloning and characterization of two human skeletal muscle alpha-actinin genes located on chromosomes 1 and 11. *Journal of Biological Chemistry,* 267*(13)*, 9281-9288.

Beighton P, Paepe AD, Steinmann B, Tsipouras P & Wenstrup RJ (1998). Ehlers-Danlos syndromes: revised nosology, Villefranche, 1997. *American Journal of Medical Genetics,* 77*(1)*, 31-37.

Bell RD, Shultz SJ, Wideman L & Henrich VC (2012a). Collagen gene variants previously associated with anterior cruciate ligament injury risk are also associated with joint laxity. *Sports Health: A Multidisciplinary Approach,* 4*(4)*, 312-318.

Bell RD, Winkler EA, Singh I, Sagare AP, Deane R, Wu Z, Holtzman DM, Betsholtz C, Armulik A & Sallstrom J (2012b). Apolipoprotein E controls cerebrovascular integrity via cyclophilin A. *Nature,* 485*(7399)*, 512-516.

Bell W, Colley J, Evans W, Darlington S & Cooper S-M (2012c). ACTN3 genotypes of Rugby Union players: distribution, power output and body composition. *Annals of Human Biology,* 39*(1)*, 19-27.

Bell W, Colley J, Gwynne J, Glazier P, Evans WD & Darlington SE (2010). ACE ID genotype and leg power in Rugby Union players. *Journal of Sports Medicine and Physical Fitness,* 50*(3)*, 350.

Bell W, Colley J, Gwynne J, Llewellyn L, Darlington S & Evans W (2009). Polymorphisms of the angiotensin-converting enzyme (ACE/ID) and their distribution in developing young adult Rugby Union players. In *Contemporary Sport, Leisure and Ergonomics*, ed. Reilly T & Atkinson G, pp. 161. Routledge, NY.

Bellefroid EJ, Kobbe A, Gruss P, Pieler T, Gurdon JB & Papalopulu N (1998). Xiro3 encodes a Xenopus homolog of the Drosophila Iroquois genes and functions in neural specification. *The EMBO Journal,* 17*(1)*, 191-203.

Ben-Zaken S, Meckel Y, Nemet D & Eliakim A (2015). Genetic score of power-speed and endurance track and field athletes. *Scandinavian Journal of Medicine & Science in Sports,* 25*(2)*, 166-174.

Benjamini Y & Hochberg Y (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society Series B (Methodological),* 57*(1)*, 289-300.

Benjamini Y & Yekutieli D (2001). The control of the false discovery rate in multiple testing under dependency. *Annals of Statistics*, 1165-1188.

Bennett ER, Reuter-Rice K & Laskowitz DT (2016). Genetic Influences in Traumatic Brain Injury. In *Translational Research in Traumatic Brain Injury*, ed. Laskovwi D & Grant G. Taylor and Francis, Boca Raton , FL.

Bennett RE, Esparza TJ, Lewis HA, Kim E, Mac Donald CL, Sullivan PM & Brody DL (2013). Human Apolipoprotein E4 Worsens Acute Axonal Pathology but Not Amyloid-β Immunoreactivity After Traumatic Brain Injury in 3× TG-AD Mice. *Journal of Neuropathology & Experimental Neurology,* 72*(5)*, 396-403.

Berk BC, Vekshtein V, Gordon HM & Tsuda T (1989). Angiotensin II-stimulated protein synthesis in cultured vascular smooth muscle cells. *Hypertension,* 13*(4)*, 305-314.

Berman Y & North KN (2010). A gene for speed: the emerging role of α-actinin-3 in muscle metabolism. *Physiology,* 25*(4)*, 250-259.

Bertuzzi R, Pasqua LA, Bueno S, Lima-Silva AE, Matsuda M, Marquezini M & Saldiva PH (2014). Is the COL5A1 rs12722 gene polymorphism associated with running economy? *PloS One,* 9*(9)*, e106581.

Birk DE, Fitch J, Babiarz J, Doane K & Linsenmayer T (1990). Collagen fibrillogenesis in vitro: interaction of types I and V collagen regulates fibril diameter. *Journal of Cell Science,* 95*(4)*, 649-657.

Blanchard A, Ohanian V & Critchley D (1989). The structure and function of α-actinin. *Journal of Muscle Research and Cell Motility,* 10*(4)*, 280-289.

Bonferroni CE (1936). Teoria Statistica delle Classi e Calcolo delle Probabilit. *Pubblicazioni del R Istituto Superiore di Scienze Economiche e Commerciali di Firenze,* 8, 3-62.

Boraita A, de la Rosa A, Heras ME, de la Torre AI, Canda A, Rabadan M, Diaz AE, Gonzalez C, Lopez M & Hernandez M (2010). Cardiovascular adaptation, functional capacity and Angiotensin-converting enzyme I/D polymorphism in elite athletes. *Revista Esoanola De Cardiologia,* 63*(7)*, 810-819.

Bottinelli R, Canepari M, Pellegrino M & Reggiani C (1996). Force-velocity properties of human skeletal muscle fibres: myosin heavy chain isoform and temperature dependence. *The Journal of Physiology,* 495*(2)*, 573-586.

Bouchard C (2015). Exercise genomics—a paradigm shift is needed: a commentary. *British Journal of Sports Medicine*, bjsports-2015-095294.

Bouchard C, An P, Rice T, Skinner JS, Wilmore JH, Gagnon J, Pérusse L, Leon AS & Rao D (1999). Familial aggregation of V˙o2max response to exercise training: results from the HERITAGE Family Study. *Journal of Applied Physiology,* 87*(3)*, 1003-1008.

Bouchard C, Boulay MR, Simoneau J-A, Lortie G & Pérusse L (1988). Heredity and Trainability of Aerobic and Anaerobic Performances an Update. *Sports Medicine,* 5*(2)*, 69-73.

Bouchard C & Perusse L (1988). Heredity and body fat. *Annual Review of Nutrition,* 8*(1)*, 259-277.

Bouchard C, Sarzynski MA, Rice TK, Kraus WE, Church TS, Sung YJ, Rao D & Rankinen T (2011). Genomic predictors of the maximal O2 uptake response to standardized exercise training programs. *Journal of Applied Physiology,* 110*(5)*, 1160-1170.

Bouchard C, Simoneau J, Lortie G, Boulay M, Marcotte M & Thibault M (1986). Genetic effects in human skeletal muscle fiber type distribution and enzyme activities. *Canadian Journal of Physiology and Pharmacology,* 64*(9)*, 1245-1251.

Bradley PS, Carling C, Diaz AG, Hood P, Barnes C, Ade J, Boddy M, Krustrup P & Mohr M (2013). Match performance and physical capacity of players in the top three competitive standards of English professional soccer. *Human Movement Science,* 32*(4)*, 808-821.

Bradley WJ, Cavanagh BP, Douglas W, Donovan TF, Morton JP & Close GL (2015). Quantification of training load, energy intake, and physiological adaptations during a rugby preseason: A case study from an elite eropean rugby union squad. *The Journal of Strength & Conditioning Research,* 29*(2)*, 534-544.

Brooks JH & Kemp S (2011). Injury-prevention priorities according to playing position in professional rugby union players. *British Journal of Sports Medicine,* 45*(10)*, 765-775.

Brooks JH & Kemp SP (2008). Recent trends in rugby union injuries. *Clinics in Sports Medicine,* 27*(1)*, 51-73.

Broos S, Malisoux L, Theisen D, Francaux M, Deldicque L & Thomis MA (2012). Role of alpha-actinin-3 in contractile properties of human single muscle fibers: a case series study in paraplegics. *PloS One,* 7*(11)*, e49281.

Broos S, Malisoux L, Theisen D, Van Thienen R, Ramaekers M, Jamart C, Deldicque L, Thomis MA & Francaux M (2016). Evidence for ACTN3 as a speed gene in isolated human muscle fibers. *PloS One,* 11*(3)*, e0150594.

Busjahn CA, Schulz-Menger J, Abdel-Aty H, Rudolph A, Jordan J, Luft FC & Busjahn A (2009). Heritability of left ventricular and papillary muscle heart size: a twin study with cardiac magnetic resonance imaging. *European Heart Journal,* 30*(13)*, 1643-1647.

Cahill N, Lamb K, Worsfold P, Headey R & Murray S (2013). The movement characteristics of English Premiership rugby union players. *Journal of Sports Sciences,* 31*(3)*, 229-237.

Calderwood C, Murray AD & Stewart W (2015). Turning people into couch potatoes is not the cure for sports concussion. *British Journal of Sports Medicine,* 0, 1-2.

Calvo M, Rodas G, Vallejo M, Estruch A, Arcas A, Javierre C, Viscor G & Ventura J (2002). Heritability of explosive power and anaerobic capacity in humans. *European Journal of Applied Physiology,* 86*(3)*, 218-225.

Cappa F, Caridi G, Gimelli G & Ghiggeri GM (1995). A new biallelic DNA polymorphism of the human COL5A1 gene. *Human Genetics,* 95*(5)*, 599-600.

Chan S, Seto JT, Houweling PJ, Yang N, North KN & Head SI (2011). Properties of extensor digitorum longus muscle and skinned fibers from adult and aged male and female Actn3 knockout mice. *Muscle Nerve,* 43*(1)*, 37-48.

Chanut-Delalande H, Bonod-Bidaud C, Cogne S, Malbouyres M, Ramirez F, Fichard A & Ruggiero F (2004). Development of a functional skin matrix requires deposition of collagen V heterotrimers. *Molecular and Cellular Biology,* 24*(13)*, 6049-6057.

Chavalarias D, Wallach JD, Li AHT & Ioannidis JP (2016). Evolution of reporting P values in the biomedical literature, 1990-2015. *Journal of the American Medical Assocition,* 315*(11)*, 1141-1148.

Chen H, Cohen P & Chen S (2010). How big is a big odds ratio? Interpreting the magnitudes of odds ratios in epidemiological studies. *Comun Stats-Simulat Comp,* 39*(4)*, 860-864.

Chen J-A, Huang Y-P, Mazzoni EO, Tan GC, Zavadil J & Wichterle H (2011). Mir-17-3p controls spinal neural progenitor patterning by regulating Olig2/Irx3 cross-repressive loop. *Neuron,* 69*(4)*, 721-735.

Chen L, Sun X, Jiang Y & Kuai L (2015). APOEε4 increases trauma induced early apoptosis via reducing delayed rectifier K+ currents in neuronal/glial co-cultures model. *Experimental Cell Research,* 334*(2)*, 294-300.

Chiu L-L, Chen T-W, Hsieh SS & Hsieh L-L (2012). ACE I/D, ACTN3 R577X, PPARD T294C and PPARGC1A Gly482Ser polymorphisms and physical fitness in Taiwanese late adolescent girls. *The Journal of Physiological Sciences,* 62*(2)*, 115-121.

Chtourou H, Chaouachi A, Hammouda O, Chamari K & Souissi N (2012). Listening to music affects diurnal variation in muscle power output. *International Journal of Sports Medicine,* 33*(1)*, 43-47.

Cieszczyk P, Krupecki K, Maciejewska A & Sawczuk M (2009). The angiotensin converting enzyme gene I/D polymorphism in Polish rowers. *International Journal of Sports Medicine,* 30*(8)*, 624-627.

Cieszczyk P, Maciejewska A, Sawczuk M, Ficek K, Eider J & Jascaniene N (2010). The angiotensin converting enzyme gene I/D polymorphism in ellite Polish and Lithuanian judo players. *Biology of Sport,* 27*(2)*, 119.

Clarke GM, Anderson CA, Pettersson FH, Cardon LR, Morris AP & Zondervan KT (2011). Basic statistical analysis in genetic case-control studies. *Nature Protocols,* 6*(2)*, 121-133.

Coelho DB, Pimenta EM, Rosse IC, Carvalho MRS, Costa GG & Silami-Garcia E (2014). Comparison of the genotypes and allele frequencies of the ACTN3 of football players from different categories. *R Bras Ci Mov,* 22*(4)*, 97-106.

Collins M, Mokone G, September A, Van der Merwe L & Schwellnus M (2009). The COL5A1 genotype is associated with range of motion measurements. *Scand J Med Sci Spor,* 19*(6)*, 803-810.

Collins M, September AV & Posthumus M (2015). Biological variation in musculoskeletal injuries: current knowledge, future research and practical implications. *British Journal of Sports Medicine,* 49*(23)*, 1497-1503.

Corella D, Arnett DK, Tucker KL, Kabagambe EK, Tsai M, Parnell LD, Lai C-Q, Lee Y-C, Warodomwichit D & Hopkins PN (2011). A high intake of saturated fatty acids strengthens the association between the fat mass and obesity-associated gene and BMI. *The Journal of Nutrition,* 141*(12)*, 2219-2225.

Crawford F, Vanderploeg R, Freeman M, Singh S, Waisman M, Michaels L, Abdullah L, Warden D, Lipsky R & Salazar A (2002). APOE genotype influences acquisition and recall following traumatic brain injury. *Neurology,* 58*(7)*, 1115-1118.

Crewther B, Kilduff L, Cook C, Cunningham D, Bunce P, Bracken R & Gaviglio C (2012). Scaling strength and power for body mass differences in rugby union players. *The Journal of Sports Medicine and Physical Fitness,* 52*(1)*, 27-32.

Crewther BT, Lowe T, Weatherby RP, Gill N & Keogh J (2009). Neuromuscular performance of elite rugby union players and relationships with salivary hormones. *The Journal of Strength & Conditioning Research,* 23*(7)*, 2046-2053.

Cross M, Kemp S, Smith A, Trewartha G & Stokes K (2015). Professional Rugby Union players have a 60% greater risk of time loss injury after concussion: a

2-season prospective study of clinical outcomes. *British Journal of Sports Medicine,* 0, 1-6.

Cross MR, Brughelli M, Brown SR, Samozino P, Gill ND, Cronin JB & Morin J-B (2014). Mechanical properties of sprinting in elite rugby union and rugby league. *International Journal of Sports Physiology & Performance,* 10*(6)*.

Cunniffe B, Proctor W, Baker JS & Davies B (2009). An evaluation of the physiological demands of elite rugby union using global positioning system tracking software. *The Journal of Strength & Conditioning Research,* 23*(4)*, 1195-1203.

Danser AJ, Schalekamp MA, Bax WA, van den Brink AM, Saxena PR, Riegger GA & Schunkert H (1995). Angiotensin-converting enzyme in the human heart effect of the deletion/insertion polymorphism. *Circulation,* 92*(6)*, 1387-1388.

De Moor MH, Spector TD, Cherkas LF, Falchi M, Hottenga JJ, Boomsma DI & De Geus EJ (2007). Genome-wide linkage scan for athlete status in 700 British female DZ twin pairs. *Twin Research and Human Genetics,* 10*(06)*, 812-820.

Decq P, Gault N, Blandeau M, Kerdraon T, Berkal M, ElHelou A, Dusfour B & Peyrin J-C (2016). Long-term consequences of recurrent sports concussion. *Acta Neurochirurgica,* 158*(2)*, 289-300.

Deutsch M, Kearney G & Rehrer N (2007). Time–motion analysis of professional rugby union players during match-play. *Journal of Sports Sciences,* 25*(4)*, 461-472.

Di Mauro M, Izzicupo P, Santarelli F, Falone S, Pennelli A, Amicarelli F, Calafiore AM, Di Baldassarre A & Gallina S (2010). ACE and AGTR1 polymorphisms and left ventricular hypertrophy in endurance athletes. *Medicine and Science in Sports and Exercise,* 42*(5)*, 915-921.

Dietze GJ & Henriksen EJ (2008). Angiotensin-converting enzyme in skeletal muscle: sentinel of blood pressure control and glucose homeostasis. *Journal of Renin-Angiotensin-Aldosterone System,* 9*(2)*, 75-88.

Djarova T, Ramakoaba A, Basson A, Mugandani S, Grace J, Cloete J & Watson G (2011a). Angiotensin-converting enzyme genotypes relationship with blood pressure, C-reactive protein and selected physical tests in Zulu South African cricketers. *African Journal of Biochemistry Research,* 5*(7)*, 197-205.

Djarova T, Watson G, Basson A, Grace J, Cloete J & Ramakoaba A (2011b). ACTN3 and TNF gene polymorphism association with C-reactive protein, uric acid, lactate and physical characteristics in young African cricket players. *African Journal of Biochemistry Research,* 5*(1)*, 22-27.

Drozdovska SB, Dosenko VE, Ahmetov II & Ilyin VN (2013). The association of gene polymorphisms with athlete status in Ukrainians. *Biology of Sport,* 30*(3)*, 163-167.

Druzhevskaya AM, Ahmetov II, Astratenkova IV & Rogozkin VA (2008). Association of the ACTN3 R577X polymorphism with power athlete status in Russians. *European Journal of Applied Physiology,* 103*(6)*, 631-634.

Dudoit S & Van Der Laan MJ (2007). *Multiple Testing Procedures with Applications to Genomics*. Springer Science & Business Media, NY.

Dzau VJ (1988a). Circulating versus local renin-angiotensin system in cardiovascular homeostasis. *Circulation,* 77*(6)*, I3-14.

Dzau VJ (1988b). Tissue renin-angiotensin system: physiologic and pharmacologic implications. *Circulation,* 77*(6)*, I1-I3.

Egorova ES, Borisova AV, Mustafina LJ, Arkhipova AA, Gabbasov RT, Druzhevskaya AM, Astratenkova IV & Ahmetov II (2014). The polygenic profile of Russian football players. *Journal of Sports Sciences,* 32*(13)*, 1286-1293.

Eider J, Cieszczyk P, Ficek K, Leonska-Duniec A, Sawczuk M, Maciejewska-Karlowska A & Zarebska A (2013). The association between D allele of the ACE gene and power performance in Polish elite athletes. *Science & Sports,* 28*(6)*, 325-330.

Eider J, Ficek K, Kaczmarczyk M, Maciejewska-Karłowska A, Sawczuk M & Cięszczyk P (2014). Endothelial nitric oxide synthase g894t (rs1799983) gene polymorphism in polish athletes. *Open Life Sciences,* 9*(3)*, 260-267.

ENCODE Project Consortium (2012). An integrated encyclopedia of DNA elements in the human genome. *Nature,* 489*(7414)*, 57-74.

England Professional Rugby Injury Surveillance Project Steering Group (2016). 2014-2015 Season report. In *England Professional Rugby Injury Surveillance Project*. Rugby Football Union, Twickenham, UK.

England Rugby (2016). Positions guide, ed. Rugby Football Union. Rugby Football Union, *http://www.englandrugby.com/*.

Erdös E & Skidgel R (1987). The angiotensin I converting enzyme. *Nature,* 56, 345-348.

Erskine R, Williams A, Jones D, Stewart C & Degens H (2014). The individual and combined influence of ACE and ACTN3 genotypes on muscle phenotypes before and after strength training. *Scandinavian Journal of Medicine & Science in Sports,* 24*(4)*, 642-648.

Eynon N, Alves AJ, Meckel Y, Yamin C, Ayalon M, Sagiv M & Sagiv M (2010). Is the interaction between HIF1A P582S and ACTN3 R577X determinant for power/sprint performance? *Metabolism,* 59*(6)*, 861-865.

Eynon N, Banting LK, Ruiz JR, Cieszczyk P, Dyatlov DA, Maciejewska-Karlowska A, Sawczuk M, Pushkarev VP, Kulikov LM & Pushkarev ED (2014). ACTN3 R577X polymorphism and team-sport performance: a study involving three European cohorts. *Journal of Science and Medicine in Sport,* 17*(1)*, 102-106.

Eynon N, Hanson ED, Lucia A, Houweling PJ, Garton F, North KN & Bishop DJ (2013a). Genes for elite power and sprint performance: ACTN3 leads the way. *Sports Medicine,* 43*(9)*, 803-817.

Eynon N, Meckel Y, Alves AJ, Nemet D & Eliakim A (2011a). Is there an interaction between BDKRB2− 9/+ 9 and GNB3 C825T polymorphisms and elite athletic performance? *Scandinavian Journal of Medicine & Science in Sports,* 21*(6)*, e242-e246.

Eynon N, Nasibulina ES, Banting LK, Cieszczyk P, Maciejewska-Karlowska A, Sawczuk M, Bondareva EA, Shagimardanova RR, Raz M & Sharon Y (2013b). The FTO A/T polymorphism and elite athletic performance: a study involving three groups of European athletes. *PloS One,* 8*(4)*, e60570.

Eynon N, Oliveira J, Meckel Y, Sagiv M, Yamin C, Sagiv M, Amir R & Duarte JA (2009). The guanine nucleotide binding protein β polypeptide 3 gene C825T polymorphism is associated with elite endurance athletes. *Experimental Physiology,* 94*(3)*, 344-349.

Eynon N, Ruiz J, Yvert T, Santiago C, Gómez-Gallego F, Lucia A & Birk R (2012). The C allele in NOS3-786 T/C polymorphism is associated with elite soccer player's status. *International Journal of Sports Medicine,* 33*(7)*, 521-524.

Eynon N, Ruiz JR, Meckel Y, Morán M & Lucia A (2011b). Mitochondrial biogenesis related endurance genotype score and sports performance in athletes. *Mitochondrion,* 11*(1)*, 64-69.

Eynon N, Ruiz JR, Oliveira J, Duarte JA, Birk R & Lucia A (2011c). Genes and elite athletes: a roadmap for future research. *The Journal of Physiology,* 589*(13)*, 3063-3070.

Fatini C, Guazzelli R, Manetti P, Battaglini B, Gensini F, Vono R, Toncelli L, Zilli P, Capalbo A & Abbate R (2000). RAS genes influence exercise-induced left ventricular hypertrophy: an elite athletes study. *Medicine and Science in Sports and Exercise,* 32*(11)*, 1868-1872.

Fedotovskaya ON, Mustafina LJ, Popov DV, Vinogradova OL & Ahmetov II (2014). A common polymorphism of the MCT1 gene and athletic performance. *International Journal of Sports Physiology & Performance,* 9*(1)*, 173-180.

Ficek K, Cieszczyk P, Kaczmarczyk M, Maciejewska-Karłowska A, Sawczuk M, Cholewinski J, Leonska-Duniec A, Stepien-Slodkowska M, Zarebska A & Stepto NK (2013). Gene variants within the COL1A1 gene are associated with reduced anterior cruciate ligament injury in professional soccer players. *Journal of Science and Medicine in Sport,* 16*(5)*, 396-400.

Ficek K, Stepien-Slodkowska M, Kaczmarczyk M, Maciejewska-Karlowska A, Sawczuk M, Cholewinski J, Leonska-Duniec A, Zarebska A, Cieszczyk P & Zmijewski P (2014). Does the A9285G Polymorphism in Collagen Type XII α1 gene associate with the risk of anterior cruciate ligament ruptures? *Balkan Journal of Medical Genetics,* 17*(1)*, 41.

Folland J, Leach B, Little T, Hawker K, Myerson S, Montgomery H & Jones D (2000). Angiotensin-converting enzyme genotype affects the response of human skeletal muscle to functional overload. *Experimental Physiology,* 85*(05)*, 575-579.

Foster B, Morse CI, Onambele-Pearson G, Ahmetov II & Williams AG (2012). Genetic variation, protein composition and potential influences on tendon properties in humans. *Open Sports Med J,* 6, 8-21.

Foster BP, Morse CI, Onambele GL & Williams AG (2014). Human COL5A1 rs12722 gene polymorphism and tendon properties in vivo in an asymptomatic population. *European Journal of Applied Physiology,* 114*(7)*, 1393-1402.

Fraas MR, Coughlan GF, Hart EC & McCarthy C (2014). Concussion history and reporting rates in elite Irish rugby union players. *Physical Therapy in Sport,* 15*(3)*, 136-142.

Frayling TM, Timpson NJ, Weedon MN, Zeggini E, Freathy RM, Lindgren CM, Perry JR, Elliott KS, Lango H & Rayner NW (2007). A common variant in the FTO gene is associated with body mass index and predisposes to childhood and adult obesity. *Science,* 316*(5826)*, 889-894.

Frazer KA, Ballinger DG, Cox DR, Hinds DA, Stuve LL, Gibbs RA, Belmont JW, Boudreau A, Hardenbol P & Leal SM (2007). A second generation human haplotype map of over 3.1 million SNPs. *Nature,* 449*(7164)*, 851-861.

Frederiksen H, Gaist D, Christian Petersen H, Hjelmborg J, McGue M, Vaupel JW & Christensen K (2002). Hand grip strength: A phenotype suitable for identifying genetic variants affecting mid-and late-life physical functioning. *Genetic Epidemiology,* 23*(2)*, 110-122.

Fuller C, Laborde F, Leather R & Molloy MG (2008). International rugby board rugby world cup 2007 injury surveillance study. *British Journal of Sports Medicine,* 42*(6)*, 452-459.

Fuller CW, Molloy MG, Bagate C, Bahr R, Brooks JH, Donson H, Kemp SP, McCrory P, McIntosh AS & Meeuwisse WH (2007). Consensus statement on injury definitions and data collection procedures for studies of injuries in rugby union. *British Journal of Sports Medicine,* 41*(5)*, 328-331.

Fuller CW, Taylor A, Kemp SP & Raftery M (2016). Rugby World Cup 2015: World Rugby injury surveillance study. *British Journal of Sports Medicine*, In press.

Fuller CW, Taylor A & Raftery M (2015a). Epidemiology of concussion in men's elite Rugby-7s (Sevens World Series) and Rugby-15s (Rugby World Cup, Junior World Championship and Rugby Trophy, Pacific Nations Cup and English Premiership). *British Journal of Sports Medicine,* 49*(7)*, 478-483.

Fuller CW, Taylor AE, Brooks JH & Kemp SP (2013). Changes in the stature, body mass and age of English professional rugby players: A 10-year review. *Journal of Sports Sciences,* 31*(7)*, 795-802.

Fuller GW, Kemp SP & Decq P (2015b). The International Rugby Board (IRB) Pitch side concussion assessment trial: a pilot test accuracy study. *British Journal of Sports Medicine,* 49*(8)*, 529-535.

Gabbett TJ (2007). Physiological and anthropometric characteristics of elite women rugby league players. *The Journal of Strength & Conditioning Research,* 21*(3)*, 875-881.

Gabbett TJ, Jenkins DG & Abernethy B (2012). Physical demands of professional rugby league training and competition using microtechnology. *Journal of Science and Medicine in Sport,* 15*(1)*, 80-86.

Garatachea N, Verde Z, Santos-Lozano A, Yvert T, Rodriguez-Romo G, Sarasa FJ, Hernández-Sánchez S, Santiago C & Lucia A (2014). ACTN3 R577X polymorphism and explosive leg-muscle power in elite basketball players. *Int J Sport Physiol Perform,* 9*(2)*.

Gardner A, Iverson G, Stanwell P, Moore T, Ellis J & Levi C (2016). A Video Analysis of Use of the New 'Concussion Interchange Rule'in the National Rugby League. *International Journal of Sports Medicine*.

Gardner A, Iverson GL, Levi CR, Schofield PW, Kay-Lambkin F, Kohler RM & Stanwell P (2015a). A systematic review of concussion in rugby league. *British Journal of Sports Medicine,* 49*(8)*, 495-498.

Gardner AJ, Iverson GL, Quinn TN, Makdissi M, Levi CR, Shultz SR, Wright DK & Stanwell P (2015b). A preliminary video analysis of concussion in the National Rugby League. *Brain injury,* 29*(10)*, 1182-1185.

Gardner AJ, Iverson GL, Williams WH, Baker S & Stanwell P (2014). A systematic review and meta-analysis of concussion in rugby union. *Sports Medicine,* 44*(12)*, 1717-1731.

Garton FC & North KN (2016). The Effect of Heterozygosity for the ACTN3 Null Allele on Human Muscle Performance. *Medicine and Science in Sports and Exercise,* 48*(3)*, 509-520.

Gauderman WJ (2002). Sample size requirements for association studies of gene-gene interaction. *American Journal of Epidemiology,* 155*(5)*, 478-484.

Gaunt TR, Rodríguez S & Day IN (2007). Cubic exact solutions for the estimation of pairwise haplotype frequencies: implications for linkage disequilibrium analyses and a web tool'CubeX'. *BMC Bioinformatics,* 8*(1)*, 428.

Gayagay G, Yu B, Hambly B, Boston T, Hahn A, Celermajer DS & Trent RJ (1998). Elite endurance athletes and the ACE I allele–the role of genes in athletic performance. *Human Genetics,* 103*(1)*, 48-50.

Geddes J, Vowles G, Nicoll J & Revesz T (1999). Neuronal cytoskeletal changes are an early consequence of repetitive head injury. *Acta Neuropathologica,* 98*(2)*, 171-178.

Geisterfer A, Peach MJ & Owens GK (1988). Angiotensin II induces hypertrophy, not hyperplasia, of cultured rat aortic smooth muscle cells. *Circulation Research,* 62*(4)*, 749-756.

George KP, Gates PE, Whyte G, Fenoglio RA & Lea R (1999). Echocardiographic examination of cardiac structure and function in elite cross trained male and female Alpine skiers. *British Journal of Sports Medicine,* 33*(2)*, 93-98.

Gilliver S, Degens H, Rittweger J, Sargeant A & Jones D (2009). Variation in the determinants of power of chemically skinned human muscle fibres. *Experimental Physiology,* 94*(10)*, 1070-1078.

Ginevičienė V, Jakaitienė A, Pranculis A, Milašius K, Tubelis L & Utkus A (2014). AMPD1 rs17602729 is associated with physical performance of sprint and power in elite Lithuanian athletes. *BMC Genetics,* 15*(1)*, 1.

Gineviciene V, Jakaitiene A, Tubelis L & Kucinskas V (2014). Variation in the ACE, PPARGC1A and PPARA genes in Lithuanian football players. *European Journal of Sport science,* 14*(sup1)*, S289-S295.

Ginevičienė V, Kučinskas V & Kasnauskienė J (2009). The angiotensin converting enzyme gene insertion/deletion polymorphism in Lithuanian professional athletes. *Acta Medica Lituanica,* 16*(1)*, 11-16.

Ginevičienė V, Pranckevičienė E, Milašius K & Kučinskas V (2010). Relating fitness phenotypes to genotypes in Lithuanian elite athletes. *Acta Medica Lituanica,* 17*(1-2)*, 1-10.

Ginevičienė V, Pranckevičienė E, Milašius K & Kučinskas V (2011). Gene variants related to the power performance of the Lithuanian athletes. *Open Life Sciences,* 6*(1)*, 48-57.

Glaister M, Howatson G, Abraham C, Lockey R, Goodwin J, Foley P & McInnes G (2008). Caffeine supplementation and multiple sprint running performance. *Medicine and Science in Sports and Exercise,* 40*(10)*, 1835.

Glenn KL, Du Z-Q, Eisenmann JC & Rothschild MF (2009). An alternative method for genotyping of the ACE I/D polymorphism. *Molecular Biology Reports,* 36*(6)*, 1305-1310.

Goh K, Chew K, Koh A, Guan M, Wong Y & Sum C (2009). The relationship between ACE gene ID polymorphism and aerobic capacity in Asian rugby players. *Singapore Medical Journal,* 50*(10)*, 997-1003.

Gokhale S & Laskowitz DT (2013). ApoE and outcome after traumatic brain injury. *Clinical Lipidology,* 8*(5)*, 561-571.

Goldman SM, Tanner CM, Oakes D, Bhudhikanok GS, Gupta A & Langston JW (2006). Head injury and Parkinson's disease risk in twins. *Annals of Neurology,* 60*(1)*, 65-72.

Gómez-Gallego F, Ruiz JR, Buxens A, Altmäe S, Artieda M, Santiago C, González-Freire M, Verde Z, Arteta D & Martínez A (2010). Are elite endurance athletes genetically predisposed to lower disease risk? *Physiological Genomics,* 41*(1)*, 82-90.

Gomez-Gallego F, Santiago C, Gonzalez-Freire M, Muniesa C, Fernández DVM, Perez M, Foster C & Lucia A (2009). Endurance performance: genes or gene combinations? *International Journal of Sports Medicine,* 30*(1)*, 66-72.

Gordon SE, Davis BS, Carlson CJ & Booth FW (2001). ANG II is required for optimal overload-induced skeletal muscle hypertrophy. *American Journal of Physiology-Endocrinology and Metabolism,* 280*(1)*, E150-E159.

Grealy R, Herruer J, Smith CL, Hiller D, Haseler LJ & Griffiths LR (2015). Evaluation of a 7-Gene Genetic Profile for Athletic Endurance Phenotype in Ironman Championship Triathletes. *PloS One,* 10*(12)*, e0145171.

Gronek P, Holdys J, Konarski J, Kryściak J & Wolc A (2013). ACE genotype in professional field hockey players. *Trends in Sport Science,* 1*(20)*, 36-40.

Grunnet LG, Nilsson E, Ling C, Hansen T, Pedersen O, Groop L, Vaag A & Poulsen P (2009). Regulation and function of FTO mRNA expression in human skeletal muscle and subcutaneous adipose tissue. *Diabetes,* 58*(10)*, 2402-2408.

Gülyaşar T, Öztürk L, Sipahi T, Bayraktar B, Metin G, Yücesir İ & Süt N (2014). GNB3 gene c. 825C>T polymorphism and performance parameters in professional basketball players. *Acta Physiologica Hungarica,* 101*(2)*, 176-184.

Hagberg JM, Ferrell RE, Katzel LI, Dengel DR, Sorkin JD & Goldberg AP (1999). Apolipoprotein E genotype and exercise training-induced increases in plasma high-density lipoprotein (HDL)- and HDL2-cholesterol levels in overweight men. *Metabolism,* 48*(8)*, 943-945.

Hakim A, Cherkas L, Spector T & MacGregor A (2003). Genetic associations between frozen shoulder and tennis elbow: a female twin study. *Rheumatology,* 42*(6)*, 739-742.

Hammond CJ, Webster AR, Snieder H, Bird AC, Gilbert CE & Spector TD (2002). Genetic influence on early age-related maculopathy: a twin study. *Ophthalmology,* 109*(4)*, 730-736.

Hardy GH (1908). Mendelian proportions in a mixed population. *Science,* 28, 49-50.

Hartman RE, Laurer H, Longhi L, Bales KR, Paul SM, McIntosh TK & Holtzman DM (2002). Apolipoprotein E4 influences amyloid deposition but not cell loss after traumatic brain injury in a mouse model of Alzheimer's disease. *The Journal of Neuroscience,* 22*(23)*, 10083-10087.

He D, Fu M, Miao S, Hotta K, Chandak GR & Xi B (2014). FTO gene variant and risk of hypertension: A meta-analysis of 57,464 hypertensive cases and 41,256 controls. *Metabolism,* 63*(5)*, 633-639.

Heffernan SM, Kilduff LP, Erskine RM, Day SH, McPhee JS, McMahon GE, Stebbings GK, Neale JP, Lockey SJ & Ribbans WJ (2016). Association of ACTN3 R577X but not ACE I/D gene variants with elite rugby union player status and playing position. *Physiological Genomics,* 48*(3)*, 196-201.

Herskind AM, McGue M, Holm NV, Sörensen TI, Harvald B & Vaupel JW (1996). The heritability of human longevity: a population-based study of 2872 Danish twin pairs born 1870–1900. *Human Genetics,* 97*(3)*, 319-323.

Heshmat H, Azab A & El-Tantawy A (2010). ACE and ACE Gene Polymorphism and Some Cardiac Parameters of Fencing, Basketball and Endurance Players. *World Journal of Sport Sciences,* 3*(2)*, 100-106.

Hidding U, Mielke K, Waetzig V, Brecht S, Hanisch U, Behrens A, Wagner E & Herdegen T (2002). The c-Jun N-terminal kinases in cerebral microglia: immunological functions in the brain. *Biochemical Pharmacology,* 64*(5)*, 781-788.

Hildebrand K, Frank C & Hart D (2004). Gene intervention in ligament and tendon: current status, challenges, future directions. *Gene Therapy,* 11*(4)*, 368-378.

Hixson JE & Vernier DT (1990). Restriction isotyping of human apolipoprotein E by gene amplification and cleavage with HhaI. *Journal of Lipid Research,* 31*(3)*, 545-548.

Hogarth MW, Garton FC, Houweling PJ, Tukiainen T, Lek M, Macarthur DG, Seto JT, Quinlan KG, Yang N & Head SI (2016). Analysis of the ACTN3 heterozygous genotype suggests that α-actinin-3 controls sarcomeric composition and muscle function in a dose-dependent fashion. *Human Molecular Genetics,* 25*(5)*, 866-877.

Holdys J, Gronek P, Krysciak J & Stanislawski D (2013). Genetic variants of uncoupling proteins-2 and-3 in relation to maximal oxygen uptake in different sports. *Acta Biochimica Polonica,* 60, 71-75.

Holdys J, Kryściak J, Stanisławski D & Gronek P (2011). ACE I/D gene polymorphism in athletes of various sports disciplines. *Human Movement,* 12*(3)*, 223-231.

Hong EP & Park JW (2012). Sample size and statistical power calculation in genetic association studies. *Genomics & Informatics,* 10*(2)*, 117-122.

Huang T, Qi Q, Li Y, Hu FB, Bray GA, Sacks FM, Williamson DA & Qi L (2014). FTO genotype, dietary protein, and change in appetite: the Preventing Overweight Using Novel Dietary Strategies trial. *The American Journal of Clinical Nutrition,* 99*(5)*, 1126-1130.

Huang Y & Mahley RW (2014). Apolipoprotein E: structure and function in lipid metabolism, neurobiology, and Alzheimer's diseases. *Neurobiology of Disease,* 72, 3-12.

Huebbe P, Nebel A, Siegert S, Moehring J, Boesch-Saadatmandi C, Most E, Pallauf J, Egert S, Muller MJ, Schreiber S, Nothlings U & Rimbach G (2011). APOE epsilon4 is associated with higher vitamin D levels in targeted replacement mice and humans. *FASEB Journal,* 25*(9)*, 3262-3270.

Hughes DC, Day SH, Ahmetov II & Williams AG (2011). Genetics of muscle strength and power: polygenic profile similarity limits skeletal muscle performance. *Journal of Sports Sciences,* 29*(13)*, 1425-1434.

Imamura Y, Scott IC & Greenspan DS (2000). The Pro-α3 (V) Collagen Chain complete primary structure, expression domains in adult and developing tissues, and comparison to the structures and expression domains of the other types V and XI procollagen chains. *Journal of Biological Chemistry,* 275*(12)*, 8749-8759.

International HapMap Consortium (2005). A haplotype map of the human genome. *Nature,* 437*(7063)*, 1299-1320.

International Rugby Board (2004). Strategic Plan. International Rugby Board, worldrugby.com.

Ioannidis JP (2005). Why most published research findings are false. *PLoS Medicine,* 2*(8)*, e124.

Ishigai Y, Mori T, Ikeda T, Fukuzawa A & Shibano T (1997). Role of bradykinin-NO pathway in prevention of cardiac hypertrophy by ACE inhibitor in rat cardiomyocytes. *American Journal of Physiology-Heart and Circulatory Physiology,* 273*(6)*, H2659-H2663.

Iwao-Koizumi K, Ota T, Hayashida M, Yonetani Y, Nakata K, Kinoshita K, Koyanagi Y & Murata S (2014). The ACTN3 Gene is a Potential Biomarker for the Risk of Non-Contact Sports Injury in Female Athletes. *Journal of Molecular Biomarkers & Diagnosis,* S6, 1-2.

Jack H (1922). Brothers in Rugby Foorball: More evidence That brothers are selected to play in the same or similar positions on football reams. *Journal of Heredity,* 13*(4)*, 161-162.

Jacobsson JA, Schiöth H & Fredriksson R (2012). The impact of intronic single nucleotide polymorphisms and ethnic diversity for studies on the obesity gene FTO. *Obesity Reviews,* 13*(12)*, 1096-1109.

Jang Y & Kim SM (2012). Influences of the G2350A polymorphism in the ACE gene on cardiac structure and function of ball game players. *Journal of Negative Results in Biomedicine,* 11, 6.

Jardine M, Wiggins T, Myburgh K & Noakes T (1988). Physiological characteristics of rugby players including muscle glycogen content and muscle fibre composition. *The South African Medical Journal,* 73*(9)*, 529-532.

Jess T, Zimmermann E, Kring SII, Berentzen T, Holst C, Toubro S, Astrup A, Hansen T, Pedersen O & Sørensen TI (2008). Impact on weight dynamics and general growth of the common FTO rs9939609: a longitudinal Danish cohort study. *International Journal of Obesity,* 32*(9)*, 1388-1394.

Jin H-J, Hwang I-W, Kim K-C, Cho H-I, Park T-H, Shin Y-A, Lee H-S, Hwang J-H, Kim A-R & Lee K-H (2015). Is there a relationship between PPARD T294C/PPARGC1A Gly482Ser variations and physical endurance performance in the Korean population? *Genes & Genomics,* 38*(4)*, 389-395.

Johnston RD, Gabbett TJ & Jenkins DG (2014). Applied sport science of rugby league. *Sports Medicine,* 44*(8)*, 1087-1100.

Jones MR, West DJ, Crewther BT, Cook CJ & Kilduff LP (2015). Quantifying positional and temporal movement patterns in professional rugby union using global positioning system. *European Journal of Sport Science,* 15*(6)*, 488-496.

Jones MR, West DJ, Harrington BJ, Cook CJ, Bracken RM, Shearer DA & Kilduff LP (2014). Match play performance characteristics that predict post-match creatine kinase responses in professional rugby union players. *BMC Sports Science, Medicine and Rehabilitation,* 6*(1)*, 1.

Jonsson JR, Game PA, Head RJ & Frewin DB (1994). The expression and localisation of the angiotensin-converting enzyme mRNA in human adipose tissue. *Blood Pressure,* 3*(1-2)*, 72-75.

Jorde LB & Wooding SP (2004). Genetic variation, classification and'race'. *Nature Genetics,* 36*(11)*, S28-S33.

Juffer P, Furrer R, Gonzalez-Freire M, Santiago C, Verde Z, Serratosa L, Morate F, Rubio J, Martin M & Ruiz J (2009). Genotype distributions in top-level soccer players: a role for ACE? *International Journal of Sports Medicine,* 30*(5)*, 387-392.

Kang BY, Kang CY, Oh SD, Bae JS, Kim KT & Lee KO (2003). The protein polymorphism of haptoglobin in Korean elite athletes. *Medical Principles and Practice,* 12*(3)*, 151-155.

Kelley K & Preacher KJ (2012). On effect size. *Psychological Methods,* 17*(2)*, 137.

Kemp S, Patricios J & Raftery M (2016). Is the content and duration of the graduated return to play protocol after concussion demanding enough? A challenge for Berlin 2016. *British Journal of Sports Medicine,* 50*(11)*, 644-645.

Kikuchi N, Min S-k, Ueda D, Igawa S & Nakazato K (2012). Higher frequency of the ACTN3 R allele and ACE DD genotype in Japanese elite wrestlers. *The Journal of Strength & Conditioning Research,* 26*(12)*, 3275-3280.

Kikuchi N, Nakazato K, Min S-k, Ueda D & Igawa S (2014a). The ACTN3 R577X polymorphism is associated with muscle power in male Japanese athletes. *The Journal of Strength & Conditioning Research,* 28*(7)*, 1783-1789.

Kikuchi N, Nakazato K, Min SK, Ueda D & Igawa S (2014b). The ACTN3 R577X polymorphism is associated with muscle power in male Japanese athletes. *The Journal of Strength & Conditioning Research,* 28*(7)*, 1783-1789.

Kilpeläinen TO, Qi L, Brage S, Sharp SJ, Sonestedt E, Demerath E, Ahmad T, Mora S, Kaakinen M & Sandholt CH (2011). Physical activity attenuates the influence of FTO variants on obesity risk: a meta-analysis of 218,166 adults and 19,268 children. *PLoS Medicine,* 8*(11)*, e1001116.

Kim W, Cho HI, Kim KC, So YH & Oh JG (2011). Relationships between digit ratio (2D: 4D), ACE gene polymorphism, and physical performance in the Korean population. *Genes & Genomics,* 33*(4)*, 407-412.

King D, Gissane C, Clark T & Marshall SW (2014). The incidence of match and training injuries in rugby league: a pooled data analysis of published studies. *Int J Sports Sci Coach,* 9*(2)*, 417-432.

Kirkham EM & Weaver EM (2015). A review of multiple hypothesis testing in otolaryngology literature. *The Laryngoscope,* 125*(3)*, 599-603.

Koch W, Latz W, Eichinger M, Ganser C, Schömig A & Kastrati A (2005). Genotyping of the angiotensin I-converting enzyme gene insertion/deletion polymorphism by the TaqMan method. *Clinical Chemistry,* 51*(8)*, 1547-1549.

Kothari ST, Chheda P, Chatterjee L & Das BR (2012). Molecular analysis of genetic variation in angiotensin I-converting enzyme identifies no association with sporting ability: First report from Indian population. *Indian Journal of Human Genetics,* 18*(1)*, 62-65.

Kristman VL, Tator CH, Kreiger N, Richards D, Mainwaring L, Jaglal S, Tomlinson G & Comper P (2008). Does the Apolipoprotein E4 Allele Predispose Varsity Athletes to Concussion? A Prospective Cohort Study. *Clinical Journal of Sport Medicine,* 18*(4)*, 322-328.

Kutner KC, Erlanger DM, Tsai J, Jordan B & Relkin NR (2000). Lower cognitive performance of older football players possessing apolipoprotein E ε4. *Neurosurgery,* 47*(3)*, 651-658.

Laguette M-J, Abrahams Y, Prince S & Collins M (2011). Sequence variants within the 3′-UTR of the COL5A1 gene alters mRNA stability: implications for musculoskeletal soft tissue injuries. *Matrix Biology,* 30*(5)*, 338-345.

Lahiri DK & Nurnberger JI (1991). A rapid non-enzymatic method for the preparation of HMW DNA from blood for RFLP studies. *Nucleic Acids Research,* 19*(19)*, 5444.

Lakens D (2013). Calculating and reporting effect sizes to facilitate cumulative science: a practical primer for t-tests and ANOVAs. *Frontiers in Psychology,* 4*(863)*, 1-12.

Lakka H-M, Lakka TA, Rankinen T, Rice T, Rao D, Leon AS, Skinner JS & Bouchard C (2006). The TNF-α G-308A polymorphism is associated with C-reactive protein levels: the HERITAGE Family Study. *Vascular pharmacology,* 44*(5)*, 377-383.

Lamas NJ, Johnson-Kerner B, Roybon L, Kim YA, Garcia-Diaz A, Wichterle H & Henderson CE (2014). Neurotrophic requirements of human motor neurons defined using amplified and purified stem cell-derived cultures. *PloS One,* 9*(10)*, e110324.

Lappalainen T, Lindström J, Paananen J, Eriksson JG, Karhunen L, Tuomilehto J & Uusitupa M (2012). Association of the fat mass and obesity-associated (FTO) gene variant (rs9939609) with dietary intake in the Finnish Diabetes Prevention Study. *British Journal of Nutrition,* 108*(10)*, 1859-1865.

Laskowitz DT, Song P, Wang H, Mace B, Sullivan PM, Vitek MP & Dawson HN (2010). Traumatic brain injury exacerbates neurodegenerative pathology: improvement with an apolipoprotein E-based therapeutic. *Journal of Neurotrauma,* 27*(11)*, 1983-1995.

Laskowitz DT & Vitek MP (2007). Apolipoprotein E and neurological disease: therapeutic potential and pharmacogenomic interactions. *Pharmacogenomics,* 8*(8)*, 959-969.

Lawrence DW, Comper P, Hutchison MG & Sharma B (2015). The role of apolipoprotein E episilon (ε)-4 allele on outcome following traumatic brain injury: A systematic review. *Brain Injury,* 29*(9)*, 1018-1031.

Lear SA, Deng WQ, Pare G, Sulistyoningrum DC, Loos RJ & Devlin A (2011). Associations of the FTO rs9939609 variant with discrete body fat depots and dietary intake in a multi-ethnic cohort. *Genetics Research,* 93*(06)*, 419-426.

Li L, Bao Y, He S, Wang G, Guan Y, Ma D, Wu R, Wang P, Huang X & Tao S (2015). The Association Between Apolipoprotein E and Functional Outcome After Traumatic Brain Injury: A Meta-Analysis. *Medicine,* 94*(46)*.

Lichtman S, Seliger G, Tycko B & Marder K (2000). Apolipoprotein E and functional recovery from brain injury following postacute rehabilitation. *Neurology,* 55*(10)*, 1536-1539.

Lidor R & Ziv G (2010). Physical and physiological attributes of female volleyball players-a review. *The Journal of Strength & Conditioning Research,* 24*(7)*, 1963-1973.

Lidor R & Ziv G (2015). On-field performances of female and male field hockey players–A review. *International Journal of Performance Analysis in Sport,* 15*(1)*, 20-38.

Lim S-T, Kim C-S, Kim W-N & Min S-K (2015). The COL5A1 genotype is associated with range of motion. *Journal of Exercise Nutrition & Biochemistry,* 19*(2)*, 49.

Liu C, Mou S & Pan C (2013). The FTO gene rs9939609 polymorphism predicts risk of cardiovascular disease: a systematic review and meta-analysis. *PLoS One,* 8*(8)*, e71901.

Liu G, Zhu H, Lagou V, Gutin B, Stallmann-Jorgensen IS, Treiber FA, Dong Y & Snieder H (2010). FTO variant rs9939609 is associated with body mass index and waist circumference, but not with energy intake or physical activity in European-and African-American youth. *BMC Medical Genetics,* 11*(1)*, 57.

Livshits G, Gao F, Malkin I, Needhamsen M, Xia Y, Yuan W, Bell CG, Ward K, Liu Y & Wang J (2016). Contribution of Heritability and Epigenetic Factors to Skeletal Muscle Mass Variation in United Kingdom Twins. *The Journal of Clinical Endocrinology & Metabolism,* 101*(6)*, 2450-2459.

Loos R, Hagberg JM, Pérusse L, Roth SM, Sarzynski MA, Wolfarth B, Rankinen T & Bouchard C (2015). Advances in exercise, fitness, and performance genomics in 2014. *Medicine and Science in Sports and Exercise,* 47*(6)*, 1105-1112.

Loos RJ & Yeo GS (2014). The bigger picture of FTO the first GWAS-identified obesity gene. *Nature Reviews Endocrinology,* 10*(1)*, 51-61.

Luis DA, Aller R, Izaola O, Conde R, Sagrado MG & Primo D (2012). Evaluation of weight loss and adipocytokines levels after two hypocaloric diets with different macronutrient distribution in obese subjects with rs9939609 gene variant. *Diabetes/Metabolism Research and Reviews,* 28*(8)*, 663-668.

Lukic I (2003). MedCalc Version 7.0.0.2. Software Review. *The Croatian Medical Journal* 44*(1)*, 120-121.

Lynch JR, Tang W, Wang H, Vitek MP, Bennett ER, Sullivan PM, Warner DS & Laskowitz DT (2003). APOE genotype and an ApoE-mimetic peptide modify the systemic and central nervous system inflammatory response. *Journal of Biological Chemistry,* 278*(49)*, 48529-48533.

Ma F, Yang Y, Li X, Zhou F, Gao C, Li M & Gao L (2013). The association of sport performance with ACE and ACTN3 genetic polymorphisms: a systematic review and meta-analysis. *PloS One,* 8*(1)*, e54685.

MacArthur DG & North KN (2004). A gene for speed? The evolution and function of α-actinin-3. *Bioessays,* 26*(7)*, 786-795.

MacArthur DG & North KN (2007). ACTN3: A genetic influence on muscle function and athletic performance. *Exercise and Sport Sciences Reviews,* 35*(1)*, 30-34.

MacArthur DG, Seto JT, Chan S, Quinlan KG, Raftery JM, Turner N, Nicholson MD, Kee AJ, Hardeman EC & Gunning PW (2008). An Actn3 knockout mouse provides mechanistic insights into the association between α-actinin-3 deficiency and human athletic performance. *Human Molecular Genetics,* 17*(8)*, 1076-1086.

Maezawa I, Maeda N, Montine TJ & Montine KS (2006). Apolipoprotein E-specific innate immune response in astrocytes from targeted replacement mice. *Journal of Neuroinflammation,* 3*(1)*, 10.

Malfait F & De Paepe A (2005). Molecular genetics in classic Ehlers–Danlos syndrome. *American Journal of Medical Genetics,* 139 C*(1)*, 17-23.

Malfait F, Wenstrup RJ & De Paepe A (2010). Clinical and genetic aspects of Ehlers-Danlos syndrome, classic type. *Genetics in Medicine,* 12*(10)*, 597-605.

Maroon JC, Winkelman R, Bost J, Amos A, Mathyssek C & Miele V (2015). Chronic traumatic encephalopathy in contact sports: a systematic review of all reported pathological cases. *PloS One,* 10*(2)*, e0117338.

Marques MC, Van den Tillaar R, Gabbett TJ, Reis VM & González-Badillo JJ (2009). Physical fitness qualities of professional volleyball players: determination of positional differences. *The Journal of Strength & Conditioning Research,* 23*(4)*, 1106-1111.

Massidda M, Bachis V, Corrias L, Piras F, Scorcu M, Culigioni C, Masala D & Calò CM (2015a). ACTN3 R577X polymorphism is not associated with team sport athletic status in Italians. *Sports Medicine-Open,* 1*(1)*, 1-5.

Massidda M, Corrias L, Bachis V, Cugia P, Piras F, Scorcu M & Calò CM (2015b). Vitamin D receptor gene polymorphisms and musculoskeletal injuries in professional football players. *Experimental and Therapeutic Medicine,* 9*(5)*, 1974-1978.

Massidda M, Corrias L, Bachis V, Culigioni C, Piras F, Scorcu M & Calò C (2014a). Genetic polymorphisms and muscle injuries among Italian Soccer Players. *Annals of Sports Medicine and Research,* 1*(1)*, 1004.

Massidda M, Corrias L, Scorcu M, Vona G & Calò MC (2012). ACTN-3 and ACE genotypes in elite male Italian athletes. *Anthropological Review,* 75*(1)*, 51-59.

Massidda M, Scorcu M & Calò CM (2014b). New genetic model for predicting phenotype traits in sports. *International Journal of Sports Physiology & Performance,* 9*(3)*, 554-560.

Matsuo T, Nakata Y, Hotta K & Tanaka K (2014). The FTO genotype as a useful predictor of body weight maintenance: Initial data from a 5-year follow-up study. *Metabolism,* 63*(7)*, 912-917.

Matsuo T, Nakata Y, Murotake Y, Hotta K & Tanaka K (2012). Effects of FTO genotype on weight loss and metabolic risk factors in response to calorie restriction among Japanese women. *Obesity,* 20*(5)*, 1122-1126.

Mattsson CM, Wheeler M, Waggott D, Caleshu C & Ashley EA (2016). Sports genetics moving forward-lessons learned from medical research. *Physiological Genomics,* 48*(3)*, 175-182.

McCaffery JM, Papandonatos GD, Peter I, Huggins GS, Raynor HA, Delahanty LM, Cheskin LJ, Balasubramanyam A, Wagenknecht LE & Wing RR (2012). Obesity susceptibility loci and dietary intake in the Look AHEAD Trial. *The American Journal of Clinical Nutrition,* 95*(6)*, 1477-1486.

McCrory P, Meeuwisse WH, Aubry M, Cantu B, Dvořák J, Echemendia RJ, Engebretsen L, Johnston K, Kutcher JS & Raftery M (2013). Consensus statement on concussion in sport: the 4th International Conference on Concussion in Sport held in Zurich, November 2012. *British Journal of Sports Medicine,* 47*(5)*, 250-258.

McLaughlin J, Linsenmayer T & Birk D (1989). Type V collagen synthesis and deposition by chicken embryo corneal fibroblasts in vitro. *Journal of Cell Science,* 94*(2)*, 371-379.

Menzel K & Hilberg T (2011). Blood coagulation and fibrinolysis in healthy, untrained subjects: effects of different exercise intensities controlled by individual anaerobic threshold. *European Journal of Applied Physiology,* 111*(2)*, 253-260.

Merritt VC & Arnett PA (2016). Apolipoprotein E (APOE) ϵ4 allele is associated with increased symptom reporting following sports concussion. *Journal of the International Neuropsychological Society,* 22*(1)*, 89-94.

Miller RH, Umberger BR & Caldwell GE (2012). Limitations to maximum sprinting speed imposed by muscle mechanical properties. *Journal of Biomechanics,* 45*(6)*, 1092-1097.

Mills M, Yang N, Weinberger R, Vander Woude DL, Beggs AH, Easteal S & North K (2001). Differential expression of the actin-binding proteins, α-actinin-2 and-3, in different species: implications for the evolution of functional redundancy. *Human Molecular Genetics,* 10*(13)*, 1335-1346.

Missitzi J, Geladas N & Klissouras V (2008). Genetic variation of maximal velocity and EMG activity. *International Journal of Sports Medicine,* 29*(3)*, 177-181.

Missitzi J, Gentner R, Geladas N, Politis P, Karandreas N, Classen J & Klissouras V (2011). Plasticity in human motor cortex is in part genetically determined. *The Journal of Physiology,* 589*(2)*, 297-306.

Miyamoto-Mikami E, Murakami H, Tsuchie H, Takahashi H, Ohiwa N, Miyachi M, Kawahara T & Fuku N (2016). Lack of association between genotype score and sprint/power performance in the Japanese population. *Journal of Science and Medicine in Sport*, In press.

Mokone G, Schwellnus M, Noakes T & Collins M (2006). The COL5A1 gene and Achilles tendon pathology. *Scandinavian Journal of Medicine & Science in Sports,* 16*(1)*, 19-26.

Moleres A, Ochoa MC, Rendo-Urteaga T, Martínez-González MA, San Julián MCA, Martínez JA & Marti A (2012). Dietary fatty acid distribution modifies obesity risk linked to the rs9939609 polymorphism of the fat mass and obesity-associated gene in a Spanish case–control study of children. *British Journal of Nutrition,* 107*(04)*, 533-538.

Montgomery H, Marshall R, Hemingway H, Myerson S, Clarkson P, Dollery C, Hayward M, Holliman D, Jubb M & World M (1998). Human gene for physical performance. *Nature,* 393*(6682)*, 221-222.

Montgomery HE, Clarkson P, Dollery CM, Prasad K, Losi M-A, Hemingway H, Statters D, Jubb M, Girvain M & Varnava A (1997). Association of angiotensin-converting enzyme gene I/D polymorphism with change in left ventricular mass in response to physical training. *Circulation,* 96*(3)*, 741-747.

Morehen JC, Routledge HE, Twist C, Morton JP & Close GL (2015). Position specific differences in the anthropometric characteristics of elite European Super League rugby players. *European Journal of Sport Science,* 15*(6)*, 523-529.

Moretti EW, Morris RW, Podgoreanu M, Schwinn DA, Newman MF, Bennett E, Moulin VG, Mba UU & Laskowitz D (2005). APOE polymorphism is associated with risk of severe sepsis in surgical patients. *Critical Care in Medicine* 33*(11)*, 2521.

Morin JB, Bourdin M, Edouard P, Peyrot N, Samozino P & Lacour JR (2012). Mechanical determinants of 100-m sprint running performance. *European Journal of Applied Physiology,* 112*(11)*, 3921-3930.

Muc M, Padez C & Manco L (2015). Influence of physical activity on the association between the FTO variant rs9939609 and adiposity in young adults. *American Journal of Human Biology,* 27*(5)*, 734-738.

Müller K, Ingebrigtsen T, Wilsgaard T, Wikran G, Fagerheim T, Romner B & Waterloo K (2009). Prediction of time trends in recovery of cognitive function after mild head injury. *Neurosurgery,* 64*(4)*, 698-704.

Munzenmaier DH & Greene AS (1996). Opposing actions of angiotensin II on microvascular growth and arterial blood pressure. *Hypertension,* 27*(3)*, 760-765.

Mustafina LJ, Naumov VA, Cieszczyk P, Popov DV, Lyubaeva EV, Kostryukova ES, Fedotovskaya ON, Druzhevskaya AM, Astratenkova IV & Glotov AS (2014). AGTR2 gene polymorphism is associated with muscle fibre composition, athletic status and aerobic performance. *Experimental Physiology,* 99*(8)*, 1042-1052.

Myerson S, Hemingway H, Budget R, Martin J, Humphries S & Montgomery H (1999). Human angiotensin I-converting enzyme gene and endurance performance. *Journal of Applied Physiology,* 87*(4)*, 1313-1316.

Nakamura O, Ishii T, Ando Y, Amagai H, Oto M, Imafuji T & Tokuyama K (2002a). Potential role of vitamin D receptor gene polymorphism in determining bone phenotype in young male athletes. *Journal of Applied Physiology,* 93*(6)*, 1973-1979.

Nakamura O, Ishii T, Mankyu H, Tsubakimoto S, Nomura T & Tokuyama K (2002b). Contribution of vitamin D receptor genotypes to bone mineral density in young male athletes with different impact loading. *European Journal of Sport Science,* 2*(2)*, 1-8.

Namipashaki A, Razaghi-Moghadam Z & Ansari-Pour N (2015). The Essentiality of Reporting Hardy-Weinberg Equilibrium Calculations in Population-Based Genetic Association Studies. *Cell Journal (Yakhteh),* 17*(2)*, 187.

Nicoll JA, Roberts GW & Graham DI (1995). Apolipoprotein E ε 4 allele is associated with deposition of amyloid β-protein following head injury. *Nature Medicine,* 1*(2)*, 135-137.

Niemi A-K & Majamaa K (2005). Mitochondrial DNA and ACTN3 genotypes in Finnish elite endurance and sprint athletes. *European Journal of Human Genetics,* 13*(8)*, 965-969.

Noble WS (2009). How does multiple testing correction work? *Nature Biotechnology,* 27*(12)*, 1135-1137.

Noé E, Ferri J, Colomer C, Moliner B & Chirivella J (2010). APOE genotype and verbal memory recovery during and after emergence from post-traumatic amnesia. *Brain injury,* 24*(6)*, 886-892.

North KN, Yang N, Wattanasirichaigoon D, Mills M, Easteal S & Beggs AH (1999). A common nonsense mutation results in α-actinin-3 deficiency in the general population. *Nature Genetics,* 21*(4)*, 353-354.

Nuzzo JL, McBride JM, Dayne AM, Israetel MA, Dumke CL & Triplett NT (2010). Testing of the maximal dynamic output hypothesis in trained and untrained subjects. *The Journal of Strength & Conditioning Research,* 24*(5)*, 1269-1276.

O'Connell K, Knight H, Ficek K, Leonska-Duniec A, Maciejewska-Karlowska A, Sawczuk M, Stepien-Slodkowska M, O'Cuinneagain D, van der Merwe W & Posthumus M (2015). Interactions between collagen gene variants and risk of anterior cruciate ligament rupture. *European Journal of Sport Science,* 15*(4)*, 341-350.

O'Connell K, Posthumus M, Schwellnus MP & Collins M (2013). Collagen genes and exercise-associated muscle cramping. *Clinical Journal of Sport Medicine,* 23*(1)*, 64-69.

Oh S (2007). The distribution of I/D polymorphism in the ACE gene among Korean male elite athletes. *Journal of Sports Medicine and Physical Fitness,* 47*(2)*, 250.

Omalu B, Bailes J, Hamilton RL, Kamboh MI, Hammers J, Case M & Robert Fitzsimmons J (2011). Emerging histomorphologic phenotypes of chronic traumatic encephalopathy in American athletes. *Neurosurgery,* 69*(1)*, 173-183.

Omalu BI, DeKosky ST, Hamilton RL, Minster RL, Kamboh MI, Shakir AM & Wecht CH (2006). Chronic traumatic encephalopathy in a national football league player: part II. *Neurosurgery,* 59*(5)*, 1086-1093.

Omalu BI, DeKosky ST, Minster RL, Kamboh MI, Hamilton RL & Wecht CH (2005). Chronic traumatic encephalopathy in a national football league player. *Neurosurgery,* 57*(1)*, 128-134.

Omalu BI, Hamilton RL, Kamboh MI, DeKosky ST & Bailes J (2010). Chronic traumatic encephalopathy (CTE) in a National Football League Player: Case report and emerging medicolegal practice questions. *Journal of Forensic Nursing,* 6*(1)*, 40-46.

Orysiak J, Busko K, Mazur-RóŻycka J, Michalski R, Gajewski J, Malczewska-Lenczowska J & Sitkowski D (2015). Relationship Between ACTN3 R577X Polymorphism and Physical Abilities in Polish Athletes. *The Journal of Strength & Conditioning Research,* 29*(8)*, 2333-2339.

Owen NJ, Watkins J, Kilduff LP, Bevan HR & Bennett MA (2014). Development of a criterion method to determine peak mechanical power output in a countermovement jump. *The Journal of Strength & Conditioning Research,* 28*(6)*, 1552-1558.

Owen SM, Venter RE, Du Toit S & Kraak WJ (2015). Acceleratory match-play demands of a Super Rugby team over a competitive season. *Journal of Sports Sciences,* 33*(19)*, 2061-2069.

Papadimitriou I, Papadopoulos C, Kouvatsi A & Triantaphyllidis C (2008). The ACTN3 gene in elite Greek track and field athletes. *International Journal of Sports Medicine,* 29*(4)*, 352-355.

Papadimitriou ID, Lucia A, Pitsiladis YP, Pushkarev VP, Dyatlov DA, Orekhov EF, Artioli GG, Guilherme JPL, Lancha AH & Ginevičienė V (2016). ACTN3 R577X and ACE I/D gene variants influence performance in elite sprinters: a multi-cohort study. *BMC Genomics,* 17*(1)*, 1.

Pasqua LA, Bueno S, Artioli GG, Lancha JR AH, Matsuda M, Marquezini MV, Lima-Silva AE, Saldiva PH & Bertuzzi R (2016). Influence of ACTN3 R577X polymorphism on ventilatory thresholds related to endurance performance. *Journal of Sports Sciences,* 34*(2)*, 163-170.

Pearson-Fuhrhop KM, Kleim JA & Cramer SC (2009). Brain plasticity and genetic factors. *Topics in Stroke Rehabilitation*.

Peeters M, Thomis M, Loos R, Derom C, Fagard R, Claessens A, Vlietinck R & Beunen G (2007). Heritability of somatotype components: a multivariate analysis. *International Journal of Obesity,* 31*(8)*, 1295-1301.

Peterson CB, Bogomolov M, Benjamini Y & Sabatti C (2016). Many Phenotypes Without Many False Discoveries: Error Controlling Strategies for Multitrait Association Studies. *Genetic Epidemiology,* 40*(1)*, 45-56.

Phillips CM, Kesse-Guyot E, McManus R, Hercberg S, Lairon D, Planells R & Roche HM (2012). High Dietary Saturated Fat Intake Accentuates Obesity Risk Associated with the Fat Mass and Obesity–Associated Gene in Adults. *The Journal of Nutrition,* 142*(5)*, 824-831.

Pimenta EM, Coelho DB, Cruz IR, Morandi RF, Veneroso CE, de Azambuja Pussieldi G, Carvalho MRS, Silami-Garcia E & Fernández JADP (2012). The ACTN3 genotype in soccer players in response to acute eccentric training. *European Journal of Applied Physiology,* 112*(4)*, 1495-1503.

Pimenta EM, Coelho DB, Veneroso CE, Coelho EJB, Cruz IR, Morandi RF, Pussieldi GDA, Carvalho MR, Garcia ES & Fernández JADP (2013). Effect of ACTN3 gene on strength and endurance in soccer players. *The Journal of Strength & Conditioning Research,* 27*(12)*, 3286-3292.

Pitsiladis Y, Wang G, Wolfarth B, Scott R, Fuku N, Mikami E, He Z, Fiuza-Luces C, Eynon N & Lucia A (2013). Genomics of elite sporting performance: what little we know and necessary advances. *British Journal of Sports Medicine,* 47*(9)*, 550-555.

Pitsiladis YP, Tanaka M, Eynon N, Bouchard C, North KN, Williams AG, Collins M, Moran CN, Britton SL & Fuku N (2016). The Athlome Project Consortium: A concerted effort to discover genomic and other "OMIC" markers of athletic performance. *Physiological Genomics,* 48, 183-190.

Pocivavsek A, Burns MP & Rebeck GW (2009a). Low-density lipoprotein receptors regulate microglial inflammation through c-Jun N-terminal kinase. *Glia,* 57*(4)*, 444-453.

Pocivavsek A, Mikhailenko I, Strickland DK & Rebeck GW (2009b). Microglial low-density lipoprotein receptor-related protein 1 modulates c-Jun N-terminal kinase activation. *Journal of Neuroimmunology,* 214*(1)*, 25-32.

Ponsford J, McLaren A, Schönberger M, Burke R, Rudzki D, Olver J & Ponsford M (2011). The association between apolipoprotein E and traumatic brain injury severity and functional outcome in a rehabilitation sample. *Journal of Neurotrauma,* 28*(9)*, 1683-1692.

Posthumus M, Schwellnus MP & Collins M (2011). The COL5A1 gene: a novel marker of endurance running performance. *Medicine and Science in Sports and Exercise,* 43*(4)*, 584-589.

Posthumus M, September AV, Keegan M, O'Cuinneagain D, Van der Merwe W, Schwellnus MP & Collins M (2009a). Genetic risk factors for anterior cruciate ligament ruptures: COL1A1 gene variant. *British Journal of Sports Medicine,* 43*(5)*, 352-356.

Posthumus M, September AV, O'Cuinneagain D, van der Merwe W, Schwellnus MP & Collins M (2009b). The COL5A1 gene is associated with increased risk of anterior cruciate ligament ruptures in female participants. *The American Journal of Sports Medicine,* 37*(11)*, 2234-2240.

Proia P, Bianco A, Schiera G, Saladino P, Contrò V, Caramazza G, Traina M, Grimaldi KA, Palma A & Paoli A (2014). PPARα gene variants as predicted performance-enhancing polymorphisms in professional Italian soccer players. *Open Access Journal of Sports Medicine,* 5, 273-278.

Pruna R, Artells R, Ribas J, Montoro B, Cos F, Muñoz C, Rodas G & Maffulli N (2013). Single nucleotide polymorphisms associated with non-contact soft tissue injuries in elite professional soccer players: influence on degree of injury and recovery time. *BMC Musculoskeletal Disorders,* 14*(1)*, 1.

Pruthi N, Chandramouli B, Kuttappa TB, Rao SL, Subbakrishna D, Abraham MP, Mahadevan A & Shankar S (2010). Apolipoprotein E polymorphism and outcome after mild to moderate traumatic brain injury: a study of patient population in India. *Neurology India,* 58*(2)*, 264.

Puthucheary Z, Skipworth JR, Rawal J, Loosemore M, Van Someren K & Montgomery HE (2011). The ACE gene and human performance. *Sports Medicine,* 41*(6)*, 433-448.

Quarrie KL, Hopkins WG, Anthony MJ & Gill ND (2013). Positional demands of international rugby union: Evaluation of player actions and movements. *Journal of Science and Medicine in Sport,* 16*(4)*, 353-359.

Raftery M (2013). Concussion and chronic traumatic encephalopathy: International Rugby Board's response. *British Journal of Sports Medicine,* 48*(2)*, 79-80.

Raftery M, Kemp S, Patricios J, Makdissi M & Decq P (2016). It is time to give concussion an operational definition: a 3-step process to diagnose (or rule out) concussion within 48 h of injury: World Rugby guideline. *British Journal of Sports Medicine,* 50*(11)*, 642-643.

Raichlen DA & Alexander GE (2014). Exercise, APOE genotype, and the evolution of the human lifespan. *Trends in Neurosciences,* 37*(5)*, 247-255.

Raleigh SM (2012). Epigenetic regulation of the ACE gene might be more relevant to endurance physiology than the I/D polymorphism. *Journal of Applied Physiology,* 112*(6)*, 1082-1083.

Rampersaud E, Mitchell BD, Pollin TI, Fu M, Shen H, O'Connell JR, Ducharme JL, Hines S, Sack P & Naglieri R (2008). Physical activity and the association of common FTO gene variants with body mass index and obesity. *Archives of Internal Medicine,* 168*(16)*, 1791-1797.

Rankinen T, Perusse L, Rauramaa R, Rivera MA, Wolfarth B & Bouchard C (2001). The human gene map for performance and health-related fitness phenotypes. *Medicine and Science in Sports and Exercise,* 33*(6)*, 855-867.

Rankinen T, Rice T, Teran-Garcia M, Rao DC & Bouchard C (2010). FTO Genotype Is Associated With Exercise Training–induced Changes in Body Composition. *Obesity,* 18*(2)*, 322-326.

Rauhio A, Uusi-Rasi K, Nikkari ST, Kannus P, Sievänen H & Kunnas T (2013). Association of the FTO and ADRB2 genes with body composition and fat distribution in obese women. *Maturitas,* 76*(2)*, 165-171.

Reilly T & Borrie A (1992). Physiology applied to field hockey. *Sports Medicine,* 14*(1)*, 10-26.

Reneland R & Lithell H (1994). Angiotensin-converting enzyme in human skeletal muscle. A simple in vitro assay of activity in needle biopsy specimens. *Scandinavian Journal of Clinical and Laboratory Investigation,* 54*(2)*, 105-111.

Rigat B, Hubert C, Alhenc-Gelas F, Cambien F, Corvol P & Soubrier F (1990). An insertion/deletion polymorphism in the angiotensin I-converting enzyme gene accounting for half the variance of serum enzyme levels. *Journal of Clinical Investigation,* 86*(4)*, 1343.

Roberts SP, Trewartha G, Higgitt RJ, El-Abd J & Stokes KA (2008). The physical demands of elite English rugby union. *Journal of Sports Sciences,* 26*(8)*, 825-833.

Roozenbeek B, Maas AI & Menon DK (2013). Changing patterns in the epidemiology of traumatic brain injury. *Nature Reviews Neurology,* 9*(4)*, 231-236.

Rosskopf D, Schwahn C, Neumann F, Bornhorst A, Rimmbach C, Mischke M, Wolf S, Geissler I, Kocher T & Grabe H (2011). The growth hormone—IGF-I axis as a mediator for the association between FTO variants and body mass index: results of the Study of Health in Pomerania. *International Journal of Obesity,* 35*(3)*, 364-372.

Roth SM, Walsh S, Liu D, Metter EJ, Ferrucci L & Hurley BF (2008). The ACTN3 R577X nonsense allele is under-represented in elite-level strength athletes. *European Journal of Human Genetics,* 16*(3)*, 391-394.

Rubanovich A & Khromov-Borisov N (2014). Theoretical analysis of the predictability indices of the binary genetic tests. *Russian Journal of Genetics: Applied Research,* 4*(2)*, 146-158.

Ruiz J, Eynon N, Meckel Y, Fiuza-Luces C, Santiago C, Gómez-Gallego F, Oliveira J & Lucia A (2011a). GNB3 C825T Polymorphism and elite athletic status: A

replication study with two ethnic groups. *International Journal of Sports Medicine,* 32*(2)*, 151-153.

Ruiz JR, Arteta D, Buxens A, Artieda M, Gómez-Gallego F, Santiago C, Yvert T, Morán M & Lucia A (2010). Can we identify a power-oriented polygenic profile? *Journal of Applied Physiology,* 108*(3)*, 561-566.

Ruiz JR, Fernández del Valle M, Verde Z, Díez-Vega I, Santiago C, Yvert T, Rodríguez-Romo G, Gómez-Gallego F, Molina J & Lucia A (2011b). ACTN3 R577X polymorphism does not influence explosive leg muscle power in elite volleyball players. *Scandinavian Journal of Medicine & Science in Sports,* 21*(6)*, e34-e41.

Saber-Ayad MM, Nassar YS & Latif IA (2013). Angiotensin-converting enzyme I/D gene polymorphism affects early cardiac response to professional training in young footballers. *Journal of Renin-Angiotensin-Aldosterone System,* 15*(3)*, 236-242.

Sabo T, Lomnitski L, Nyska A, Beni S, Maronpot R, Shohami E, Roses A & Michaelson D (2000). Susceptibility of transgenic mice expressing human apolipoprotein E to closed head injury: the allele E3 is neuroprotective whereas E4 increases fatalities. *Neuroscience,* 101*(4)*, 879-884.

Sachidanandam R, Weissman D, Schmidt SC, Kakol JM, Stein LD, Marth G, Sherry S, Mullikin JC, Mortimore BJ & Willey DL (2001). A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. *Nature,* 409*(6822)*, 928-933.

Sadoshima J-i, Xu Y, Slayter HS & Izumo S (1993). Autocrine release of angiotensin II mediates stretch-induced hypertrophy of cardiac myocytes in vitro. *Cell,* 75*(5)*, 977-984.

Salanti G, Amountza G, Ntzani EE & Ioannidis JP (2005). Hardy–Weinberg equilibrium in genetic association studies: an empirical evaluation of reporting, deviations, and power. *European Journal of Human Genetics,* 13*(7)*, 840-848.

Salles JI, Amaral MV, Aguiar DP, Lira DA, Quinelato V, Bonato LL, Duarte MEL, Vieira AR & Casado PL (2015). BMP4 and FGF3 haplotypes increase the risk of tendinopathy in volleyball athletes. *Journal of Science and Medicine in Sport,* 18*(2)*, 150-155.

Santiago C, González-Freire M, Serratosa L, Morate FJ, Meyer T, Gómez-Gallego F & Lucia A (2008). ACTN3 genotype in professional soccer players. *British Journal of Sports Medicine,* 42*(1)*, 71-73.

Santiago C, Ruiz J, Muniesa C, González-Freire M, Gómez-Gallego F & Lucia A (2010). Does the polygenic profile determine the potential for becoming a world-class athlete? Insights from the sport of rowing. *Scandinavian Journal of Medicine & Science in Sports,* 20*(1)*, e188-e194.

Santiago C, Ruiz JR, Rodríguez-Romo G, Fiuza-Luces C, Yvert T, Gonzalez-Freire M, Gómez-Gallego F, Morán M & Lucia A (2011). The K153R polymorphism in the myostatin gene and muscle power phenotypes in young, non-athletic men. *PloS One,* 6*(1)*, e16323.

Saunders C, September A, Xenophontos S, Cariolou M, Anastassiades L, Noakes T & Collins M (2007). No association of the ACTN3 gene R577X polymorphism

with endurance performance in Ironman Triathlons. *Annals of Human Genetics,* 71*(6)*, 777-781.

Savage J, Hooke C, Orchard J & Parkinson R (2013). The incidence of concussion in a professional Australian rugby league team, 1998–2012. *Journal of Sports Medicine,* 2013.

Sawczuk M, Timshina YI, Astratenkova IV, Maciejewska-Karlowska A, Leonska-Duniec A, Ficek K, Mustafina LJ, Cieszczyk P, Klocek T & Ahmetov, II (2013). The -9/+9 polymorphism of the bradykinin receptor Beta 2 gene and athlete status: a study involving two European cohorts. *Human Biology,* 85*(5)*, 741-756.

Scanlan AT, Tucker PS & Dalbo VJ (2014). A comparison of linear speed, closed-skill agility, and open-skill agility qualities between backcourt and frontcourt adult semiprofessional male basketball players. *The Journal of Strength & Conditioning Research,* 28*(5)*, 1319-1327.

Schuelke M, Wagner KR, Stolz LE, Hübner C, Riebel T, Kömen W, Braun T, Tobin JF & Lee S-J (2004). Myostatin mutation associated with gross muscle hypertrophy in a child. *New England Journal of Medicine,* 350*(26)*, 2682-2688.

Schulz TJ, Huang TL, Tran TT, Zhang H, Townsend KL, Shadrach JL, Cerletti M, McDougall LE, Giorgadze N & Tchkonia T (2011). Identification of inducible brown adipocyte progenitors residing in skeletal muscle and white fat. *Proceedings of the National Academy of Sciences,* 108*(1)*, 143-148.

Schutte NM, Nederend I, Hudziak JJ, Bartels M & de Geus EJ (2016). A Twin-Sibling Study and Meta-Analysis on the Heritability of Maximal Oxygen Consumption. *Physiological Genomics,* 48, 210-219.

Schwellnus MP, Thomson A, Derman W, Jordaan E, Readhead C, Collins R, Morris I, Van der Linde E & Williams A (2014). More than 50% of players sustained a time-loss injury (> 1 day of lost training or playing time) during the 2012 Super Rugby Union Tournament: a prospective cohort study of 17 340 player-hours. *British Journal of Sports Medicine,* 48*(17)*, 1306-1315.

Scott AC, Roe N, Coats AJ & Piepoli MF (2003). Aerobic exercise physiology in a professional rugby union team. *International Journal of Cardiology,* 87*(2)*, 173-177.

Sedeaud A, Marc A, Schipman J, Tafflet M, Hager J-P & Toussaint J-F (2012). How they won Rugby World Cup through height, mass and collective experience. *British Journal of Sports Medicine,* 46, 580-584.

Sedeaud A, Vidalin H, Tafflet M, Marc A & Toussaint J (2013). Rugby morphologies:"Bigger and taller", reflects an early directional selection. *The Journal of Sports Medicine and Physical Fitness,* 53*(2)*, 185-191.

Sen S & Burmeister M (2008). Hardy-Weinberg analysis of a large set of published association studies reveals genotyping error and a deficit of heterozygotes across multiple loci. *Human Genomics,* 3*(1)*, 36-52.

September AV, Cook J, Handley CJ, van der Merwe L, Schwellnus MP & Collins M (2009). Variants within the COL5A1 gene are associated with Achilles tendinopathy in two populations. *British Journal of Sports Medicine,* 43*(5)*, 357-365.

Seto JT, Chan S, Turner N, MacArthur DG, Raftery JM, Berman YD, Quinlan KG, Cooney GJ, Head S & Yang N (2011). The effect of α-actinin-3 deficiency on muscle aging. *Experimental Gerontology,* 46*(4)*, 292-302.

Seto JT, Quinlan KG, Lek M, Zheng XF, Garton F, MacArthur DG, Hogarth MW, Houweling PJ, Gregorevic P & Turner N (2013). ACTN3 genotype influences muscle performance through the regulation of calcineurin signaling. *Journal of Cinical Iinvestigation,* 123*(10)*, 4255-4263.

Sgourou A, Fotopoulos V, Kontos V, Patrinos GP & Papachatzopoulou A (2012). Association of genome variations in the renin-angiotensin system with physical performance. *Human Genomics,* 6*(1)*, 1-7.

Sharples AP, Hughes DC, Deane CS, Saini A, Selman C & Stewart CE (2015). Longevity and skeletal muscle mass: the role of IGF signalling, the sirtuins, dietary restriction and protein intake. *Aging Cell,* 14*(4)*, 511-523.

Silventoinen K, Magnusson PK, Tynelius P, Kaprio J & Rasmussen F (2008). Heritability of body size and muscle strength in young adulthood: a study of one million Swedish men. *Genetic Epidemiology,* 32*(4)*, 341-349.

Simoneau J-A & Bouchard C (1995). Genetic determinism of fiber type proportion in human skeletal muscle. *The FASEB Journal,* 9*(11)*, 1091-1095.

Smart D, Hopkins WG, Quarrie KL & Gill N (2014). The relationship between physical fitness and game behaviours in rugby union players. *European Journal of Sport Science,* 14*(supp 1)*, S8-S17.

Smart DJ, Hopkins WG & Gill ND (2013). Differences and changes in the physical characteristics of professional and amateur rugby union players. *The Journal of Strength & Conditioning Research,* 27*(11)*, 3033-3044.

Smemo S, Tena JJ, Kim K-H, Gamazon ER, Sakabe NJ, Gómez-Marín C, Aneas I, Credidio FL, Sobreira DR & Wasserman NF (2014). Obesity-associated variants within FTO form long-range functional connections with IRX3. *Nature,* 507*(7492)*, 371-375.

Sonestedt E, Gullberg B, Ericson U, Wirfält E, Hedblad B & Orho-Melander M (2011). Association between fat intake, physical activity and mortality depending on genetic variation in FTO. *International Journal of Obesity,* 35*(8)*, 1041-1049.

Sonestedt E, Roos C, Gullberg B, Ericson U, Wirfält E & Orho-Melander M (2009). Fat and carbohydrate intake modify the association between genetic variation in the FTO genotype and obesity. *The American Journal of Clinical Nutrition,* 90*(5)*, 1418-1425.

Spencer M, Lawrence S, Rechichi C, Bishop D, Dawson B & Goodman C (2004). Time–motion analysis of elite field hockey, with special reference to repeated-sprint activity. *Journal of Sports Sciences,* 22*(9)*, 843-850.

Stewart W, McNamara P, Lawlor B, Hutchinson S & Farrell M (2015). Chronic traumatic encephalopathy: a potential late and under recognized consequence of rugby union? *QJM: Journal of the Association of Physicians,* 109*(1)*, 11-15.

Stølen T, Chamari K, Castagna C & Wisløff U (2005). Physiology of soccer. *Sports Medicine,* 35*(6)*, 501-536.

Summers KM, Snodgrass SJ & Callister R (2014). Predictors of calf cramping in rugby league. *The Journal of Strength & Conditioning Research,* 28*(3)*, 774-783.

Sun M, Chen S, Adams SM, Florer JB, Liu H, Kao WW-Y, Wenstrup RJ & Birk DE (2011). Collagen V is a dominant regulator of collagen fibrillogenesis: dysfunctional regulation of structure and function in a corneal-stroma-specific Col5a1-null mouse model. *Journal of Cell Science,* 124*(23)*, 4096-4105.

Taddio A, Ipp M, Thivakaran S, Jamal A, Parikh C, Smart S, Sovran J, Stephens D & Katz J (2012). Survey of the prevalence of immunization non-compliance due to needle fears in children and adults. *Vaccine,* 30*(32)*, 4807-4812.

Tanaka C, Kamide K, Takiuchi S, Miwa Y, Yoshii M, Kawano Y & Miyata T (2003). An alternative fast and convenient genotyping method for the screening of angiotensin converting enzyme gene polymorphisms. *Hypertension Research,* 26*(4)*, 301-306.

Tanaka M, Wang G & Pitsiladis YP (2016). Advancing sports and exercise genomics: moving from hypothesis-driven single study approaches to large multi-omics collaborative science. *Physiological Genomics*, Ahead of print.

Tanofsky-Kraff M, Han JC, Anandalingam K, Shomaker LB, Columbo KM, Wolkoff LE, Kozlosky M, Elliott C, Ranzenhofer LM & Roza CA (2009). The FTO gene rs9939609 obesity-risk allele and loss of control over eating. *The American Journal of Clinical Nutrition,* 90*(6)*, 1483-1488.

Taylor RR, Mamotte CD, Fallon K & van Bockxmeer FM (1999). Elite athletes and the gene for angiotensin-converting enzyme. *Journal of Applied Physiology,* 87*(3)*, 1035-1037.

Teasdale GM, Nicoll JA, Murray G & Fiddes M (1997). Association of apolipoprotein E polymorphism with outcome after head injury. *The Lancet,* 350*(9084)*, 1069-1071.

Terrell TR, Bostick RM, Abramson R, Xie D, Barfield W, Cantu R, Stanek M & Ewing T (2008). APOE, APOE promoter, and Tau genotypes and risk for concussion in college athletes. *Clinical Journal of Sport Medicine,* 18*(1)*, 10-17.

Thomaes T, Thomis M, Onkelinx S, Fagard R, Matthijs G, Buys R, Schepers D, Cornelissen V & Vanhees L (2011). A genetic predisposition score for muscular endophenotypes predicts the increase in aerobic power after training: the CAREGENE study. *BMC Genetics,* 12*(1)*, 1.

Thomaes T, Thomis M, Onkelinx S, Goetschalckx K, Fagard R, Lambrechts D & Vanhees L (2013). Genetic predisposition scores associate with muscular strength, size, and trainability. *Medicine and Science in Sports and Exercise,* 45*(8)*, 1451-1459.

Thompson PD, Tsongalis GJ, Seip RL, Bilbie C, Miles M, Zoeller R, Visich P, Gordon P, Angelopoulos TJ & Pescatello L (2004). Apolipoprotein E genotype and changes in serum lipids and maximal oxygen uptake with exercise training. *Metabolism,* 53*(2)*, 193-202.

Tiainen K, Sipilä S, Alen M, Heikkinen E, Kaprio J, Koskenvuo M, Tolvanen A, Pajala S & Rantanen T (2004). Heritability of maximal isometric muscle strength in older female twins. *Journal of Applied Physiology,* 96*(1)*, 173-180.

Tierney RT, Mansell JL, Higgins M, McDevitt JK, Toone N, Gaughan JP, Mishra A & Krynetskiy E (2010). Apolipoprotein E genotype and concussion in college athletes. *Clinical Journal of Sport Medicine,* 20*(6)*, 464-468.

Trafimow D & Marks M (2015). Editorial, basic and applied social psychology. *Basic and Applied Social Psychology,* 37*(1)*, 1-2.

Trewartha G, Preatoni E, England ME & Stokes KA (2015). Injury and biomechanical perspectives on the rugby scrum: a review of the literature. *British Journal of Sports Medicine,* 49*(7)*, 425-433.

Tucker R, Santos-Concejero J & Collins M (2013). The genetic basis for elite running performance. *British Journal of Sports Medicine,* 47*(9)*, 545-549.

Turgut G, Turgut S, Genc O, Atalay A & Atalay E (2004). The angiotensin converting enzyme I/D polymorphism in Turkish athletes and sedentary controls. *Acta Medica,* 47*(2)*, 133-136.

Twist C & Worsfold P (2015). *The science of rugby*. Routledge, Abingdon, OX.

Uhorchak JM, Scoville CR, Williams GN, Arciero RA, Pierre PS & Taylor DC (2003). Risk factors associated with noncontact injury of the anterior cruciate ligament a prospective four-year evaluation of 859 west point cadets. *The American Journal of Sports Medicine,* 31*(6)*, 831-842.

Varley I, Hughes DC, Greeves JP, Stellingwerff T, Ranson C, Fraser WD & Sale C (2015). RANK/RANKL/OPG pathway: Genetic associations with stress fracture period prevalence in elite athletes. *Bone,* 71, 131-136.

Vincent B, De Bock K, Ramaekers M, Van den Eede E, Van Leemputte M, Hespel P & Thomis MA (2007). ACTN3 (R577X) genotype is associated with fiber type distribution. *Physiological Genomics,* 32*(1)*, 58-63.

Visscher PM, Brown MA, McCarthy MI & Yang J (2012). Five years of GWAS discovery. *The American Journal of Human Genetics,* 90*(1)*, 7-24.

Visscher PM, Medland SE, Ferreira MA, Morley KI, Zhu G, Cornes BK, Montgomery GW & Martin NG (2006). Assumption-free estimation of heritability from genome-wide identity-by-descent sharing between full siblings. *PLoS Genetics,* 2*(3)*, e41.

Vogel A, Holbrook K, Steinmann B, Gitzelmann R & Byers P (1979). Abnormal collagen fibril structure in the gravis form (type I) of Ehlers-Danlos syndrome. *Laboratory Investigation* 40*(2)*, 201-206.

Voisin S, Cieszczyk P, Pushkarev VP, Dyatlov DA, Vashlyayev BF, Shumaylov VA, Maciejewska-Karlowska A, Sawczuk M, Skuza L & Jastrzebski Z (2014). EPAS1 gene variants are associated with sprint/power athletic performance in two cohorts of European athletes. *BMC Genomics,* 15*(382)*, 1-11.

Wackerhage H, Miah A, Harris RC, Montgomery HE & Williams AG (2009). Genetic research and testing in sport and exercise science: A review of the issues. *Journal of Sports Sciences,* 27*(11)*, 1109-1116.

Wåhlén K, Sjölin E & Hoffstedt J (2008). The common rs9939609 gene variant of the fat mass-and obesity-associated gene FTO is related to fat cell lipolysis. *Journal of Lipid Research,* 49*(3)*, 607-611.

Waples RS (2015). Testing for Hardy–Weinberg proportions: have we lost the plot? *Journal of Heredity,* 106*(1)*, 1-19.

Webborn N, Williams A, McNamee M, Bouchard C, Pitsiladis Y, Ahmetov I, Ashley E, Byrne N, Camporesi S & Collins M (2015). Direct-to-consumer genetic testing for predicting sports performance and talent identification: Consensus statement. *British Journal of Sports Medicine,* 49*(23)*, 1486-1491.

Weinberg W (1908). *On the demonstration of heredity in man*. Prentice Hall, Englewood Cliffs, NJ.

Weisgraber KH (1994). Apolipoprotein E: structure-function relationships. *Advances in Protein Chemistry,* 45, 249-302.

Wenstrup RJ, Florer JB, Brunskill EW, Bell SM, Chervoneva I & Birk DE (2004). Type V collagen controls the initiation of collagen fibril assembly. *Journal of Biological Chemistry,* 279*(51)*, 53331-53337.

Wenstrup RJ, Florer JB, Davidson JM, Phillips CL, Pfeiffer BJ, Menezes DW, Chervoneva I & Birk DE (2006). Murine model of the Ehlers-Danlos syndrome col5a1 haploinsufficiency disrupts collagen fibril assembly at multiple stages. *Journal of Biological Chemistry,* 281*(18)*, 12888-12895.

West D, Cunningham D, Bevan H, Crewther B, Cook C & Kilduff L (2013). Influence of active recovery on professional rugby union player's ability to harness postactivation potentiation. *The Journal of Sports Medicine and Physical Fitness,* 53*(2)*, 203-208.

Wigginton JE, Cutler DJ & Abecasis GR (2005). A note on exact tests of Hardy-Weinberg equilibrium. *The American Journal of Human Genetics,* 76*(5)*, 887-893.

Wilber RL & Pitsiladis YP (2012). Kenyan and Ethiopian distance runners: what makes them so good. *International Journal of Sports Physiolpgy and Performance,* 7*(2)*, 92-102.

Williams AG & Folland JP (2008). Similarity of polygenic profiles limits the potential for elite human physical performance. *The Journal of Physiology,* 586*(1)*, 113-121.

Williams FM, Kalson NS, Fabiane SM, Mann DA & Deehan DJ (2015a). Joint Stiffness Is Heritable and Associated with Fibrotic Conditions and Joint Replacement. *PloS One,* 10*(7)*, e0133629.

Williams S, Trewartha G, Kemp S & Stokes K (2013). A meta-analysis of injuries in senior men's professional Rugby Union. *Sports Medicine,* 43*(10)*, 1043-1055.

Williams S, Trewartha G, Kemp SP, Brooks JH, Fuller CW, Taylor AE, Cross MJ & Stokes KA (2015b). Time loss injuries compromise team success in Elite Rugby Union: a 7-year prospective study. *British Journal of Sports Medicine,* 0, 1-6.

Wing MR, Ziegler J, Langefeld CD, Ng MC, Haffner SM, Norris JM, Goodarzi MO & Bowden DW (2009). Analysis of FTO gene variants with measures of obesity and glucose homeostasis in the IRAS Family Study. *Human Genetics,* 125*(5-6)*, 615-626.

Woehning A, Schultz J, Roeder E, Moeltner A, Isermann B, Nawroth P, Wolfrum C & Rudofsky G (2013). The A-allele of the common FTO gene variant rs9939609 complicates weight maintenance in severe obese patients. *International Journal of Obesity,* 37*(1)*, 135-139.

Wolfarth B, Rankinen T, Hagberg JM, Loos RJ, Pérusse L, Roth SM, Sarzynski MA & Bouchard C (2014). Advances in exercise, fitness, and performance genomics in 2013. *Medicine and Science in Sports snd Exercise,* 46*(5)*, 851-859.

Woods D, Hickman M, Jamshidi Y, Brull D, Vassiliou V, Jones A, Humphries S & Montgomery H (2001). Elite swimmers and the D allele of the ACE I/D polymorphism. *Human Genetics,* 108*(3)*, 230-232.

World Medical Association (2013). World Medical Association Declaration of Helsinki: ethical principles for medical research involving human subjects. *Journal of the American Medical Association,* 310*(20)*, 2191.

Xu J, Turner A, Little J, Bleecker ER & Meyers DA (2002). Positive results in association studies are associated with departure from Hardy-Weinberg equilibrium: hint for genotyping error? *Human Genetics,* 111*(6)*, 573-574.

Yamak B, Yuce M, Bagci H & Imamoglu O (2015). Association between Sport Performance and Alpha-Actinin-3 Gene R577X Polymorphism. *International Journal of Human Genetics,* 15*(1)*, 13-19.

Yang N, MacArthur DG, Gulbin JP, Hahn AG, Beggs AH, Easteal S & North K (2003). ACTN3 genotype is associated with human elite athletic performance. *The American Journal of Human Genetics,* 73*(3)*, 627-631.

Yang N, MacArthur DG, Wolde B, Onywera VO, Boit MK, Lau SYM-A, Wilson RH, Scott RA & Yannis P (2007). The ACTN3 R577X polymorphism in East and West African athletes. *Medicine and Science in Sports and Exercise,* 39*(11)*, 1985–1988.

Zeng S, Jiang J-X, Xu M-H, Xu L-S, Shen G-J, Zhang A-Q & Wang X-H (2014). Prognostic value of apolipoprotein E epsilon4 allele in patients with traumatic brain injury: a meta-analysis and meta-regression. *Genetic Testing and Molecular Biomarkers,* 18*(3)*, 202-210.

Zhai G, Stankovich J, Ding C, Scott F, Cicuttini F & Jones G (2004). The genetic contribution to muscle strength, knee pain, cartilage volume, bone size, and radiographic osteoarthritis: a sibpair study. *Arthritis & Rheumatism,* 50*(3)*, 805-810.

Zhang B, Tanaka H, Shono N, Miura S, Kiyonaga A, Shindo M & Saku K (2003). The I allele of the angiotensin-converting enzyme gene is associated with an increased percentage of slow-twitch type I fibers in human skeletal muscle. *Clinical Genetics,* 63*(2)*, 139-144.

Zhang X, Qi Q, Zhang C, Smith SR, Hu FB, Sacks FM, Bray GA & Qi L (2012). FTO Genotype and 2-Year Change in Body Composition and Fat Distribution in Response to Weight-Loss Diets The POUNDS LOST Trial. *Diabetes,* 61*(11)*, 3005-3011.

Zhou W, Xu D, Peng X, Zhang Q, Jia J & Crutcher KA (2008). Meta-analysis of APOE 4 allele and outcome after traumatic brain injury. *Journal of Neurotrauma,* 25*(4)*, 279-290.

Ziegler A, Van Steen K & Wellek S (2011). Investigating Hardy–Weinberg equilibrium in case–control or cohort studies or meta-analysis. *Breast Cancer Research and Treatment,* 128*(1)*, 197-201.

Ziv G & Lidor R (2009). Physical attributes, physiological characteristics, on-court performances and nutritional strategies of female and male basketball players. *Sports Medicine,* 39*(7)*, 547-568.

Ziv G & Lidor R (2010). Vertical jump in female and male basketball players—A review of observational and experimental studies. *Journal of Science and Medicine in Sport,* 13*(3)*, 332-339.

Zoppi N, Gardella R, De Paepe A, Barlati S & Colombi M (2004). Human fibroblasts with mutations in COL5A1 and COL3A1 genes do not organize collagens and fibronectin in the extracellular matrix, down-regulate α2β1 integrin, and recruit αvβ3 instead of α5β1 integrin. *Journal of Biological Chemistry,* 279*(18)*, 18157-18168.

Zweig MH & Campbell G (1993). Receiver-operating characteristic (ROC) plots: a fundamental evaluation tool in clinical medicine. *Clinical Chemistry,* 39*(4)*, 561-577.